University of Canterbury

Ph.D Thesis

New Concepts for Operating Ring Laser Gyroscopes

Author:
Richard Graham

Supervisor:
Dr. Jon-Paul Wells
Associate Supervisor:
Prof. Roger Reeves

Submitted in Partial Fulfilment of the Requirements for the Degree
Doctor of Philosophy in Physics

December 13, 2010

Abstract

A ring laser gyroscope (gyro) is an active laser interferometer designed to sense rotation through the Sagnac frequency shift encountered by two beams travelling in opposite directions around a closed path. The classes of devices considered in this thesis are the large and ultra-large ring laser gyros. These instruments are designed for direct measurement of earth rotation rate and find applications in geodesy, geophysics, and tests of physical theories.

The research presented in this thesis focuses on the demonstration of new techniques for operating ring laser gyros. The main goal of these techniques has been the correction for variations in the geometry of an ultra-large ring laser gyro, UG-3. This instrument is a 77 m perimeter ultra-large ring laser gyro of heterolithic construction and is the primary instrument used in the experiments presented here.

UG-3 has been used to demonstrate measurement of earth strains which have been used to correct for changes in the geometry of the instrument. It has also been used to demonstrate a control technique where the co-rotating beams were alternately offset allowing the number of wavelengths around the perimeter to be counted and a Sagnac rotation signal to be obtained.

Among the most important outcomes of this research of interest to the large ring laser gyro community is that we now understand most of the problems that would affect a next generation ring laser gyro. This understanding allows us to choose an operational technique best suited to the measurements being made and thus maximise the scientific potential of the instrument. Additionally, the development of a new standard for data storage and an associated suite of software to acquire, query and analyse ring laser data is expected to improve collaboration with the wider research community.

Other research outcomes of more general interest include the analysis of how oscillation of a single mode is established in a high finesse laser cavity. We demonstrate that the ultimate mode of operation can be selected with a 'seed' beam of exceptionally low intensity. An interesting related outcome is the demonstration of Sagnac beat frequency measurement during the ring down of a ring cavity, a type of measurement immune to dispersive and flow related frequency shifts.

Contents

1. Summary

1.1. Motivation

Over the past 15 years, large ring laser gyros have progressed from demonstrating the principle of a stand-alone absolute rotation sensor to routinely making observations of earth rotation rate to an accuracy of around 1 part in 10^8 over a few hours. As a result of these improvements, large ring laser gyros have proven themselves useful instruments for making a diverse range of measurements. These include measurement of long term effects such as polar wobble [1] as well as short term measurements such as those required for observation of teleseismic waves [2]. We have good reason to expect that many more scientifically important applications lie in wait of further improvements in performance.

Obtaining the present levels of performance has largely been accomplished by pushing four key parameters. These are physical size, stability, mirror quality and the application of various corrections for geophysical phenomena. The basic technology (the helium-neon laser) and detection method (direct Sagnac beat signal detection) has remained largely unchanged.

Recent experience with the ultra-large ring laser UG-2 was significant in that it was the first instrument to show poorer performance than its smaller predecessor. As a result it seems fair to characterise the current period in the history of the large ring laser as 'the end of the beginning'; i.e. the period where performance improvements can no longer be expected as a direct result of pushing the traditional parameters, in particular size. Realising further performance improvements is likely to require a 'smarter' approach, more targeted at the specific types of measurements which we would like to make. To this end, the main goals of this research include:

1. Developing a better understanding of the physical processes limiting perfor-

mance in our instruments.

2. Demonstrating new correction techniques.
3. Providing 'proof of principle' demonstrations for novel operational techniques intended to allow a rotation signal to be extracted while avoiding the limitations inherent in the traditional operational techniques.

1.2. Thesis outline

Chapter 2 begins by explaining the basic physics of the ring laser gyro, and then more specifically describing our instruments, in particular the UG series of lasers which were used for most of the experiments presented in this thesis. Following this we outline various categories of effects which limit the performance of ring laser gyros and give specific examples. We then review the ways in which the performance of ring laser gyros are characterised, with specific examples from our instruments. Lastly we review a number of interesting operational techniques that have been applied in the past to improve the performance of ring laser gyros.

Chapter 3 begins with an overview of the helium-neon laser system presented in the context of ring laser gyros and demonstrates software that was written for calculating gain curves. Following this analysis it is clear that a number of aspects of the system require further investigation. To this end we then present results of computational and experimental investigations of the mode configurations and behaviour during startup, then experimental investigations of gas temperature and saturation intensity, followed lastly by a review of both cold-cavity and plasma related refractive effects.

Chapter 4 presents several different experiments, all of which were conceived with the goal of demonstrating techniques for obtaining signals for correcting variations in the geometry of large ring laser gyros. Firstly we outline the mathematics necessary for describing geometric variations in rectangular and triangular ring lasers. We then present observations made using *multiple corner phase tracking* and *contrast ratio monitoring*, both of which can provide correction signals. Next we consider various techniques for recovering correction signals from a ring laser running with multiple longitudinal modes. Then we present results from an experiment in which the offset between co-rotating beams was alternated regularly in order to recover correction signals. Increasing the alternation rate motivates the next section which

details experiments in *injection seeding*. The natural progression of this technique leads toward *in-ring-down* measurement. Experiments demonstrating this technique and discussion of the implications to the ultimate sensitivity of ring lasers form the final section of the chapter.

Chapter 5 details several topics related to possible applications of ring laser gyros. These topics are otherwise unrelated. The first set of topics discussed are related to geophysics. Studies of the microseismic background, of permanent surface deformations due to earthquakes, and of rotations induced by tidal forces are presented. Then we present studies of three precession effects and their potential for measurement. These are the Lense-Thirring, Thomas and de Sitter (geodedic) precessions. Lastly two potential applications related to control of the mirror suspension systems of large interferometers such as VIRGO/LIGO are presented. These include the application of a ring laser as a tilt detector, and the application of a system of ring laser gyros as sensors in an active feedback system for controlling a large mirror suspension system.

Chapter 6 begins by taking a broad view of the scientific process as it has been applied historically in ring laser research from raw data acquisition, though validation and analysis to publication. We outline specific requirements and design principles necessary for increasing the efficiency of the process. A theme underlying these requirements is the necessity for rigorous standardisation in data acquisition and storage. To this end we present an XML data specification for storing ring laser data. We show that this specification, in concert with a native XML database forms a powerful tool for identifying relationships present in ring laser data. In this chapter we also document several data acquisition techniques that have been put to use both historically and in the experiments presented elsewhere in this thesis. We present both the mathematics underlying these techniques and practical implementation details.

Chapter 7 concludes the main body of the thesis. We first summarise all the factors limiting the performance of ring laser gyros. We then summarise potential solutions to each of these problems, many of which have been demonstrated to some extent in this thesis. We present recommendations for how these individual solutions could be applied together in the construction of a next generation ring laser gyro. Taking the potential measurements such an instrument could make into account, we outline a set of features that such a device should have. We follow this up with some analysis

in which we determine a short list of optimal locations for constructing an instrument implementing the described features. Lastly we take the opportunity to summarise future experiments that could be made with current ring lasers.

In keeping with the authors philosophy to ensure that the work presented in this thesis can be easily built upon, two comprehensive appendices are provided. These are written from the point of view of providing a 'users manual' to the various tools written during the course of preparing this thesis. The first documents the software written for data analysis. The second focuses on data acquisition software. This includes a question-answer style troubleshooting guide and documentation for all of the command line utilities and LabVIEW VIs. The reader of the PDF version of this document will find the source code for the implementations of each of the documented programs attached to the associated page.

The end matter begins with lists of the figures, tables and program listings presented in the thesis. This is followed by a bibliography containing all the references used in the thesis in the IOP format. This is followed by a comprehensive glossary of terms which provides a convenient reference for the precise definitions of all terms that have been used in a specific context in the thesis. This is followed by a list of nomenclature which provides definitions for all the variables and operators which have a universal definition throughout the thesis; if the meaning of a symbol in an equation is not defined in a specific context in the surrounding text then it can be found in this list. Lastly, a comprehensive index is provided.

1.3. Acknowledgements

I would like to acknowledge the generous contributions of the following people who have made a significant contribution to my supervision or the scientific results presented in this thesis:

- Dr. Bob Hurst (UC), who has had a hand in virtually all aspects of the research presented here and provided many useful ideas, guidance and encouragement.
- Dr. Jon-Paul Wells (UC, supervisor), who has helped provide the direction for the research and guidance with day to day issues.
- Dr. Ulrich Schreiber (Geodätisches Observatorium Wettzell, Technische Uni-

versität München), who has helped with geophysical aspects of the research and provided valuable feedback relating to data acquisition and control requirements.

- Prof. Angela DiVirgilio (INFN, Università di Pisa), who acted as supervisor for the work related to gravity wave detection, completed in Pisa.
- Mr. Robert Thirkettle (UC), who provided day to day technical assistance.

I would also like to acknowledge the following people who have provided data or whose work has been built on or used indirectly in this research,

- Dr. André Gebauer (Geodätisches Observatorium Wettzell, Technische Universität München), who provided data from the G ring and feedback on various data acquisition related issues.
- Dr. Dennis Milbert (National Geodetic Survey), author of original Fortran solid earth tide calculation program.
- Mr. Mark E. Smith (Sandia National Laboratories), author of the LabVIEW XMLRPC implementation.
- Mr. Jim Kring (JKI Software) and the OpenG community.
- Mr. Wolfgang Meier, eXist-db main developer.
- The New Zealand GeoNet project team.
- The python, numpy, scipy, matplotlib and matplotlib/basemap communities, who have developed excellent open source software tools which are put to use in virtually every section of this thesis.

I would also like to acknowledge the following people whose support and help with proofreading is much appreciated:

- Ms. Trà Đinh (University of Washington)
- Dr. Ewan Orr (UC)
- Mr. Sebastian Horvath (UC)
- Mr. Nishanthan Rabeendran (UC)

1.4. Usage

[1] http://www.lconz.ac.nz
[2] http://creativecommons.org/licenses/by/3.0/nz/

2. Background

The goal of this chapter is to provide the background material which underpins the later chapters. This requires explaining the physics underlying the operation of ring laser gyros and their limitations, and detailing the operational and performance characteristics of our instruments. Given that a theme of the overall thesis is looking at new ways of operating ring laser gyros, we also review interesting alternative operational techniques that have been tried in the past.

2.1. The ring laser

French physicist Georges Sagnac was the first to realise the potential of a ring interferometer to measuring rotation. In a 1913 experiment Sagnac [3] became the first to report measurements of rotation with a turntable-mounted interferometer through what is now known as the *Sagnac effect* [4]. The effect arises when a beam of light of wavelength λ is split with each part arranged to travel at speed c in opposite directions around a closed polygonal path enclosing vector area $\mathbf{A}$. If the interferometer is subject to vector rotation rate $\mathbf{\Omega}$ in inertial space while the beams are travelling, then there will be a difference in the round trip time of the two beams

$$\delta t = \frac{4\mathbf{A} \cdot \mathbf{\Omega}}{c} \tag{2.1}$$

and a corresponding relative phase shift

$$\delta\phi = \frac{8\pi\mathbf{A} \cdot \mathbf{\Omega}}{\lambda c} \tag{2.2}$$

between the two beams. After recombining the beams this phase shift can be measured from the resulting interference fringes.

Interestingly, Sagnac's original motivation was to detect relative motion of the ether. The first derivation of the Sagnac phase shift (Equation 2.2) was ether-theoretic, though modern relativistic derivations give the same result [4].

Michelson and Gale famously measured the phase shift caused by earth rotation in their 1925 experiment [5, 6] with an interferometer of perimeter 1.9 km and area $2.08 \times 10^5\,\mathrm{m}^2$. Modern fibre-optic gyros operate on this principle.

In the case of a rotating active laser interferometer (first demonstrated by Macek and Davis [7]) an additional constraint is introduced. Here we require an integer number of wavelengths around the cavity and consequently there must be a difference in the optical frequencies of the beams proportional to the rotation rate. The beat frequency which arrises after mixing the two beams is typically in the audio range and can generally be measured to much greater precision than the corresponding phase shift from a passive device.

The rotation induced frequency splitting, or *Sagnac frequency* of a ring laser takes the form

$$f_s = \frac{4\mathbf{A}\cdot\boldsymbol{\Omega}}{\lambda P} \tag{2.3}$$

where P is the perimeter. This can be expressed alternately as

$$f_s = \frac{4A\Omega_E \sin(\phi + \theta_T)}{\lambda P} \tag{2.4}$$

for a ring laser located at latitude ϕ with θ_T angle of tilt downwards (positive) to the north and measuring earth rotation rate Ω_E.

The factor P in the denominator of Equation 2.3 and the elimination of the large factor c comes about because the mode spacing, or *free spectral range* (FSR) of a ring cavity is

$$\mathrm{FSR} = c/P. \tag{2.5}$$

Note that a phase change $\delta\phi$ of 2π has the same effect as a change in the optical frequency of FSR.

The reader interested in a full derivation of Equation 2.2 and Equation 2.3 for a general polygonal path should consult the recent publication by Hurst *et al.* [8] or the rigorous treatment based on Stokes's theorem by Jacobs and Zanoni [9].

Many effects exist which have the potential to cause a deviation from the ideal behaviour of Equation 2.3. However let us first neglect the specific cause of the deviation and simply recognise that all such effects can in general be considered

to cause an equivalent perturbation in phase $\Delta\phi$ of one of the laser beams. This perturbation then causes a change in the measured Sagnac frequency

$$\Delta f_s = \frac{c}{2\pi P}\Delta\phi. \tag{2.6}$$

If we consider the overall 'signal to noise ratio' of this type of measurement

$$\begin{aligned} s/n &= \frac{f_s}{\Delta f_s} \\ &= \frac{8\pi \mathbf{A}\cdot\boldsymbol{\Omega}}{\lambda c}\frac{1}{\Delta\phi} \end{aligned} \tag{2.7}$$

we can see that increasing the area of the laser would be expected to improve the overall performance of the device[1].

2.2. Canterbury ring lasers

In an effort to continually strive for improved performance, the ring laser group at Canterbury[2] has historically gone down the path of building ever larger devices. This has included devices of both *monolithic* and *hetrolithic* construction. Monolithic construction is where the geometry of the laser can be regarded as fixed due to construction on a base made of a single piece of high stability material (e.g. Zerodur™ in the C-II and G lasers). Conversely, hetrolithic devices have geometry fixed only by construction on a base made of multiple materials not especially regarded for their stability, for example the UG lasers which are constructed on basalt rock.

All of our current large ring lasers have some design elements in common, in particular:

- Operation with the helium-neon system at 632.8 nm using high quality multilayer dielectric super-mirrors.
- A rectangular geometry.
- The whole cavity is filled with helium-neon gas and gain happens at a narrow *gain tube* section. This makes the mirrors the only optical surfaces.
- The gain medium is excited by a radio frequency (RF) field which is capacitively coupled into the gaintube.

[1] Note that in this simple analysis it is assumed that the magnitude of $\Delta\phi$ is independent of $\mathbf{A}$.
[2] In close collaboration with a German group, based at the Geodetic Observatory at Wettzell.

- The laser power is stabilised by controlling the RF power with a servo (feedback) system.
- The output power is in the nano-Watt range.
- The co-rotating beams are combined in a 'beam combiner' prism and the resulting Sagnac signal is detected by low noise photomultiplier tubes.

Table 2.1 shows all the geometric specifications of all the recently constructed ring laser gyros. Note that here the *geometric scale factor* is just the geometric part of the scale factor in Equation 2.3, i.e. A/P.

Name	Area (m^2)	Perimeter (m)	Scale Factor (m)	f_s (Hz)	FSR (MHz)
C-II	1	4	0.25	79.6	74.948
GEO	1.8	5.4	0.33	102.6	55.517
PR-1	2.56	6.4	0.4	113.5	46.843
G-0	12.25	14	0.875	288.314	21.414
G	16	16	1	348.6	18.737
UG-1/3	366.83	76.9	4.77	1512.63	3.898
UG-2	834.34	120.435	6.93	2176.827	2.489

Table 2.1. Geometric specifications of various ring lasers

2.2.1. The UG series of lasers

The majority of the experimental work in this thesis was done with the UG series of lasers, in different configurations which we refer to as UG-1, UG-2 and UG-3. UG is an abbreviation for 'ultra gross' (German) and the lasers are representative of a class of devices we term *ultra-large ring lasers*, distinct from the *large ring lasers* with perimeter of the order 4 m to 30 m.

The UG series of lasers have several design elements in common, in particular:

- They have been constructed in the Cashmere cavern laboratory. This is the same site as several of the previous Canterbury ring lasers. This cavern was excavated from basalt lava flows on the north flank of the Canterbury port hills in New Zealand and is located at 43.574613° S, 172.623281° E. The cavern was excavated in the early years of the Second World War and was originally

intended to house a command centre, however the concrete reinforcing was only partially completed before changing fortunes in the war made the cavern unnecessary.

- The laser cavity is built from 100 mm diameter stainless steel pipes, 5 m to 6 m in length, connected by copper seals.
- The 'corner boxes' (see Figure 2.2) which house the mirrors are connected to the pipes via a system of four bellows which are designed to exert equal force when under vacuum or at atmospheric pressure. Each corner box is supported by a bearing which allows micrometer-controlled lever arms to tilt and rotate the corner box. The whole corner assembly is installed on a steel base-plate set into concrete that was poured directly onto solid rock.
- The arms are orientated 9.51° counter-clockwise from north-south and east-west lines. The plane of the laser slopes downwards 0.004° to the south and 0.424° to the east. When combined with the orientation this gives an effective latitude of -43.511°, or in Equation 2.3), $\phi + \theta_T = 46.489°$.

In the UG-1/UG-3 configuration we used 1-inch diameter mirrors made by Research Electro-Optics, Inc. with 20 m radius of curvature at all corners.

With the UG-2 configuration, it was necessary to substitute 2-inch diameter mirrors with 70 m radius of curvature at the eastern corners. This choice was made to give a stable cavity[1] and to arrange a beam profile that would not be obstructed at the mirrors or gain tube (6 mm diameter). The asymmetry in this configuration causes considerable astigmatism in the output beams.

We typically fill the lasers with 0.2 mbar of neon (either natural or equal ^{20}Ne:^{22}Ne isotope ratio) followed by helium up to a total of 3.0 mbar for UG-2 or somewhat higher (sometimes up to 8 mbar) with the UG-1/UG-3 configuration. The choice of gas parameters is discussed further in Chapter 3.

The dimensions of the UG lasers are shown in Figure 2.1. These were found with a calibrated steel tape of linear density w_0=0.11 kg m^{-1} and cross section A_0=1.4 × 10^{-5} m^2. In making these measurements the tape was tensioned to the specified T=50 N and laid across the vacuum flanges of each arm leaving spans of length l. The measure-

[1] Refer to Currie *et al.* [10] for a review of the cavity stability criteria in rectangular ring lasers.

ments were corrected for thermal expansion of the tape and the extent of the sag between the flanges, assuming a sag at the mid point of

$$\text{sag} = -\frac{1}{8}\left(\frac{1}{1+\frac{T}{A_0 E}}\right)\frac{w_0}{T}l^2 \tag{2.8}$$

[11]. Here, the elastic modulus (stress/strain) E=207 GPa for steel [11]. This was used to establish the positions of the corner boxes and in turn the mirrors. The positions of the laser beams on the mirrors were found using a straight edge adjusted to obstruct half of the beam. For UG-2, the perimeter measured by steel tape was found to be 6.5 mm shorter than expected from measurement of the FSR (see Section 3.4.1 and Section 4.5). This discrepancy was apportioned equally to the individual side lengths. For UG-1/UG-3, the measurements were obtained from an earlier survey and corrected using the un-changed west arm as a reference.

2.2.2. History of the UG lasers

The UG-1 laser was first constructed late in year 2000. Initially the laser was operated in the UG-1-air configuration, where a closed gain tube with Brewster angle windows was used and the laser cavity was formed in air. Following successful operation of this configuration, the 100 mm stainless steel pipes were installed along with an open gain tube.

In 2005 the UG-1 laser was extended to form UG-2. This involved extending the north and south arms to 39.7 m in length resulting in a rectangular configuration.

Beginning on July 2$^{\text{nd}}$ 2008 UG-2 was disassembled and the pipes returned to the UG-1 configuration. While the geometric configuration is the same as UG-1, this laser is denoted UG-3 because during the reassembly process opportunity was taken to make several upgrades. These involved installing new optical breadboards at all corners to accommodate improved beam combiner optics, and upgrading the bearings which support the corner boxes. Additionally, a laboratory which had been formally used for the C-1 laser was upgraded with a new optical table installed, an optical fibre running to the UG-3 north-west corner and signal cables to all UG corners. This laboratory is now the primary space for experiments associated with UG-3. Electrical signals can now be patched from here to any location in the Cashmere cavern via a system of patch panels.

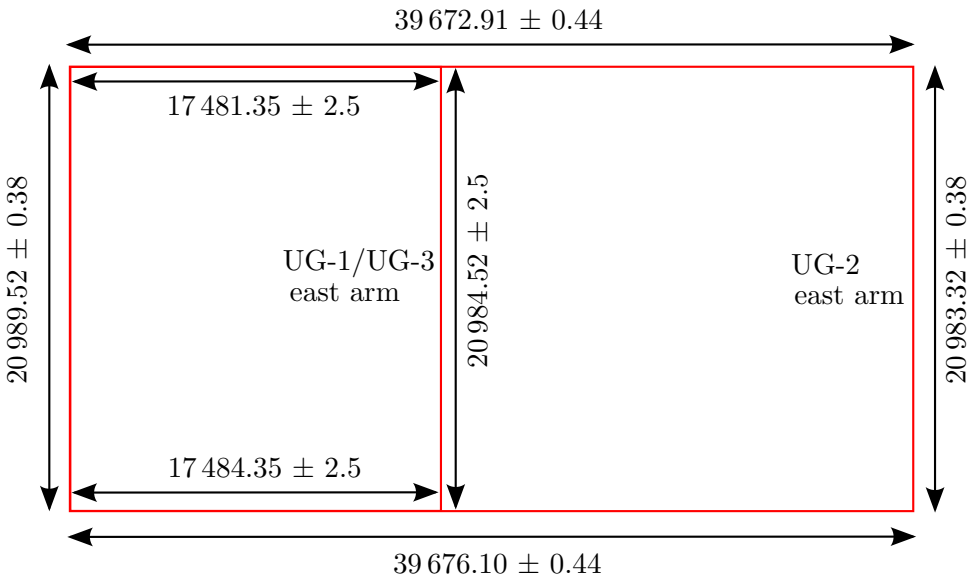


Figure 2.1. Dimensions of the UG lasers. All measurements are in milimeters.

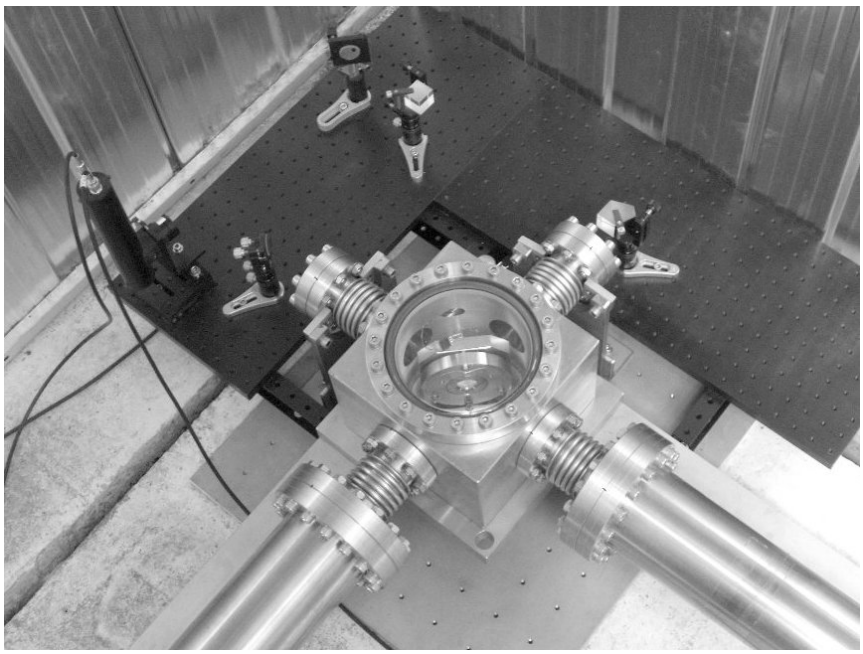


Figure 2.2. UG-3 corner box.

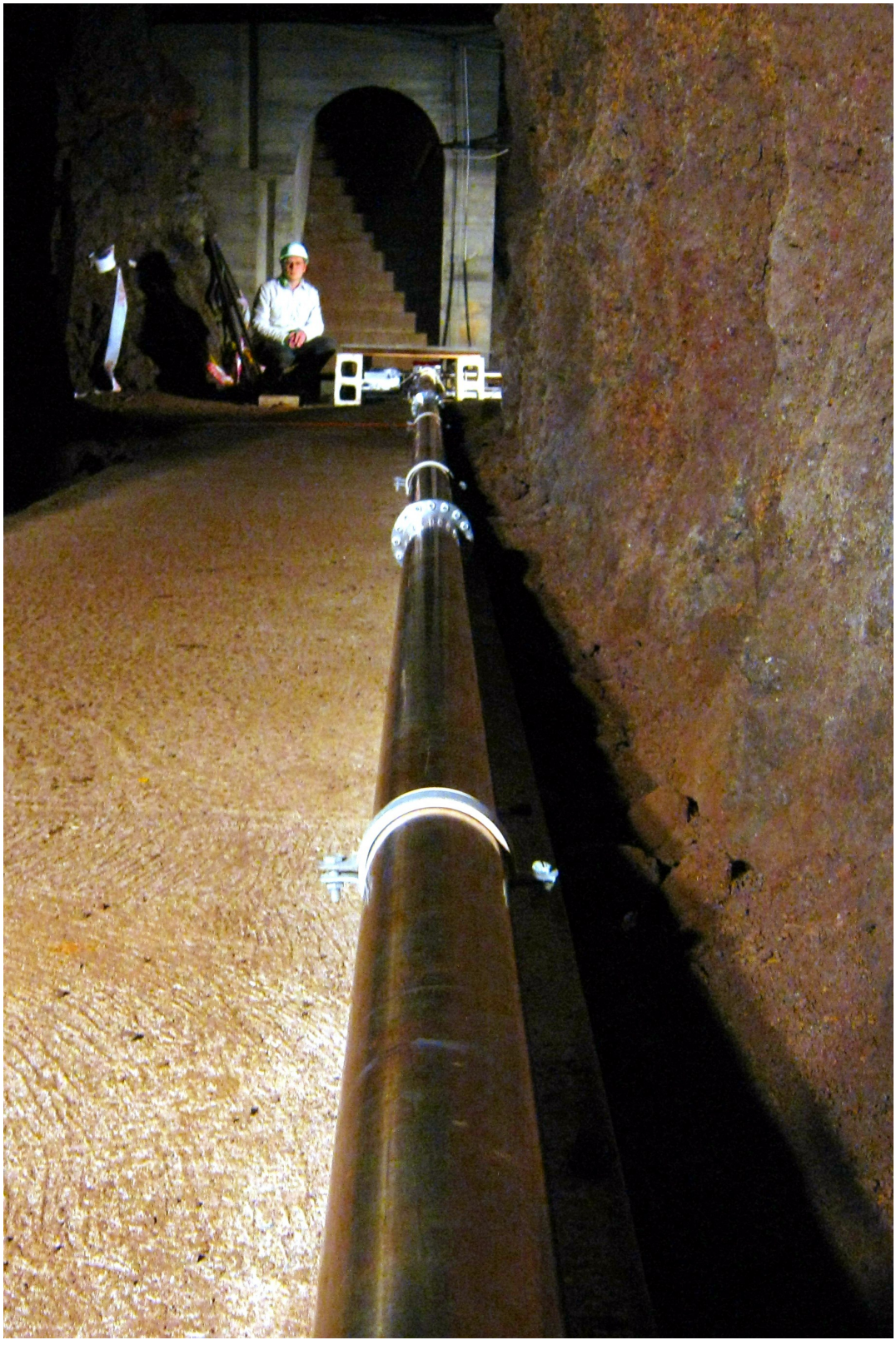


Figure 2.3. UG-3 east arm (with the author for scale).

By November 21st 2008 UG-3 was operational. This period of operation was short lived however. On December 12th 2008 the laser was extensively damaged by a large rock which fell from the cavern ceiling and broke the laser pipe on the north side. The laser lost vacuum catastrophically and in the process dirt from the cavern floor was sucked at high speed around the cavity, impacting the laser mirrors and rendering them useless for our purposes.

Following a later rockfall in the same area, it was clear that blasting was required in the region of localised instability to render the cavern safe. The north side was disassembled to accommodate this. With the blasting complete reconstruction could begin. Over several weeks the author and others cleaned and reassembled the pipes and mirror boxes. After reassembly it was found that the rate of pressure rise due to outgassing from the tubes was unacceptable. To fix this the laser pipes were baked using high temperature heating tapes while pumping on the cavity. Because of the size of the cavity and electric power limitations it was necessary to do this in sections, working around the cavity over a period of several weeks.

With the baking nearly complete, on May 15th 2009 the laser was again damaged, this time due to the failure of a ion-pump connector, possibly due to overheating. The cavity was again contaminated with air, though fortunately the mirrors suffered only minor damage and were able to be restored by cleaning. However it was necessary to repeat the baking procedure.

On July 9th 2009 UG-3 was at last restored to operational status. Over the following weeks the upgraded beam combining optics were re-installed and we were able to continue earth rotation observations.

2.3. Types of errors

There are a number of effects that can cause the response of a ring laser gyro to deviate from the ideal. These can be broadly characterised into effects which cause *null shifts*, *scale factor shifts*, *lock-in artefacts* and various classes of *random noise* sources. These are illustrated in Figure 2.4. In the following sections we will outline some examples which fit into these categories.

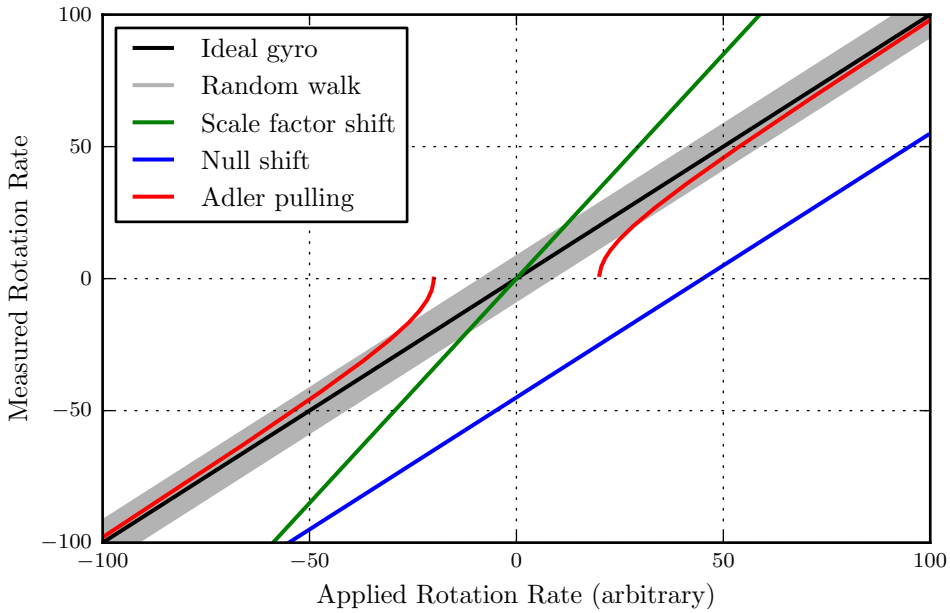


Figure 2.4. Types of errors that can cause deviations from ideal response of a ring laser gyro.

2.3.1. Null shifts

Null shifts can be modelled by simply adding an extra constant term to the right hand side of Equation 2.3.

An example of a cause of null shift is gas flow. A net flow of gas in a particular direction can cause a bias in the velocity distribution of the atoms and in turn bias the optical frequency of the counter rotating beams. The classic example is Langmuir flow occurring in DC excited plasmas. This causes a flow of gas toward the cathode along the centre of the gain tube and towards the anode along the walls [12, 13]. Aronowitz [14] however has noted that observed gas flows can not be fully explained by the relatively simple Langmuir model and that much of the work on understanding and minimising gas flow effects is empirical and proprietary information, belonging to companies who manufacture ring laser gyros for navigation. Gas flows are further explored in Section 3.8.2. In DC-excited ring lasers, gas flows are usually cancelled by operating two identical discharge tubes optically in series but with opposite polarity. In this configuration there is a severe constraint on the balance of the two currents,

of the order of 1 µA [15, 16].

Magnetic fields, applied to the helium-neon plasma have also been observed to cause null shifts [17]. Aronowitz has demonstrated an approximately linear null shift with applied magnetic field of $0.8\,^{\circ}\,\mathrm{h}^{-1}\,\mathrm{G}^{-1}$ [12], however we do not see any such effect. The mechanism here has not been thoroughly explained. We can speculate that it may be the result of Faraday rotation in the plasma (such rotations have been observed by Tobias [18]) in combination with a slight non-linear polarisation caused by mirror birefringence from imperfections in the coating.

2.3.2. Scale factor shifts

The simplest examples of scale factor shifts are the direct changes in geometry of the ring. In a hetrolithic ring laser, tidal strains and atmospheric pressure changes are major concerns. This topic is the primary concern of Chapter 4.

Tidal variations and atmospheric pressure changes can also cause changes in the angle θ_T for a ring laser attached to the earth. Schreiber *et al.* [19] have demonstrated correlation of gravitation tilt signals (from a precision tilt meter) with ring laser gyro measurements. Tilt measurements can be used to correct ring laser signals. This becomes challenging at the limiting resolution of the tilt measurements and also with heterolithic lasers where the overall tilt is not accurately defined by measurement at a single point. In addition, a tilt meter (typically a pendulum-like device) is unable to distinguish true tilts from perturbations in the local angle of the gravitational field vector and linear accelerations (such as those due to seismic events). Some perturbations to the gravitational field vector are caused by the lunar and solar gravitational fields and are thus predictable and can be corrected for in a straightforward manner. Others are the result of movements of ocean water and the atmosphere and not as easily corrected for.

Changes in the Sagnac frequency caused by changes in tilt could be minimised by building the laser so that the plane is perpendicular to the earth rotation axis. In this configuration the rate of change of Sagnac frequency with tilt would be zero. This configuration would also maximise the scale factor.

Other types of scale factor shifts exist which can be modelled by including an additional multiplication factor to the right hand side of Equation 2.3. An example of

this type of scale factor shift is the change in optical frequency of the counter rotating beams due to dispersion. Dispersion is discussed in more detail in Section 3.8.1. In brief, the magnitude of the scale factor shift due to dispersion is proportional to the single pass gain. In most examples where a helium-neon laser is operating on the 632.8 nm transition, the single pass gain is sufficiently small for the effect to be of relatively minor concern.

2.3.3. Adler Pulling and Lock in

In an ideal ring laser gyro, the counter rotating beams would occupy the same physical space but yet be completely independent. In practise the beams can be coupled due to *backscatter* from slight imperfections on the mirror surfaces. The mechanism responsible for the frequency coupling is analogous to the phase locking of two adjacent pendulum clocks via the wall motions, famously noted by Huygens in the 1600's. Adler's 1946 [20] description of the process applied to tuned circuits provides the mathematical basis for the description in relation to ring laser gyros.

The most commonly used model of the backscatter process was presented by Aronowitz [12]. In this simplified model we consider isotropic scattering particles or defects on the mirror surfaces. These cause a small fraction (r_s) of the intensity of the light incident on the mirror surfaces to be scattered. A small fraction of the scattered light (r) will be sufficiently well aligned to couple back into the counter-rotating beam, essentially mapping back onto itself within the diffraction limit. If d is the diameter of the beam at the mirror then

$$\frac{r}{r_s} = \frac{\lambda^2}{16d^2}. \tag{2.9}$$

Note that a more complete treatment would consider the size distribution of the particles and where necessary apply the *Mie scattering* (for particles of similar size to a wavelength) and Rayleigh scattering formulations. Additionally, in a more complete treatment it would be necessary to apply a correction for the matching of the scatter distributions to the transverse Hermite-Gaussian intensity profile of the counter rotating beam. This would be a very complex procedure, only amenable to numerical calculation.

Assuming however that the scattering particles and defects are smaller than a wavelength, we can apply Rayleigh scattering for a more accurate treatment. This pro-

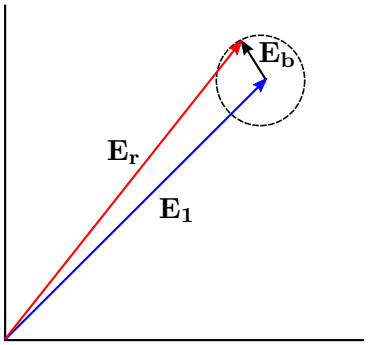


Figure 2.5. Phasor diagram illustrating backscatter pulling.

duces a dipole pattern which for our lasers (s-polarisation) has a maximum in the plane of the laser and the equivalent of Equation 2.9 becomes

$$\frac{r}{r_s} = \frac{3\lambda^2}{8\pi A} \tag{2.10}$$

with A is the $1/e^2$ spot area of the mirror. Note that this is only a statistical expectation value, the true value depends on the sum of the randomly phased amplitudes scattered from the individual particles/defects.

Based on measurement of cavity losses and typical specifications for our mirrors (transmission $\sim 0.2 \times 10^{-6}$, absorption $\sim 2 \times 10^{-6}$) we can infer a scattering intensity $r_s \sim 1 \times 10^{-5}$. The typical coupling of intensity into the counter-rotating beam is between approximately 1×10^{-12} and 1×10^{-14}.

How this scattered light causes frequency pulling can be explained with the help of the phasor diagram shown in Figure 2.5. Here the phasor $\mathbf{E_1}$ represents the amplitude of the laser beam, while the smaller phasor $\mathbf{E_b}$ represents the *backscatter phasor*. The backscatter phasor has length $\sqrt{\frac{r}{r_s}}|\mathbf{E_1}|$ (assuming clockwise and counterclockwise beams have equal power). This phasor rotates at the Sagnac frequency about the dashed circle. The resultant phasor $\mathbf{E_r}$ represents the amplitude of the 'pulled' beam, which differs in frequency relative to that represented by the phasor E_1 by

$$\Delta f = \frac{rc}{P} \sin\beta \tag{2.11}$$

where β is the difference in phase angle between $\mathbf{E_1}$ and the backscatter phasor.

As the Sagnac frequency decreases, the individual frequencies can be pulled close enough together to cause the beams to phase lock, in which case the instrument

becomes useless as a rotation sensor. The Sagnac frequency at which this occurs is known as the *lock-in threshold* f_l and is illustrated in Figure 2.4. Unless arranged with the plane of the laser parallel to the earth rotation axis, a properly operating large ring laser gyro should never be able to phase lock because the sidereal rotation rate should be more than sufficient to bias it out of the lock-in region[1]. However the frequency pulling can still introduce an error in the Sagnac frequency. The pulled Sagnac frequency is

$$f_s' = \sqrt{f_s^2 - f_l^2} \tag{2.12}$$

and the lock in threshold is given by

$$f_l = \frac{rc}{2\pi P}. \tag{2.13}$$

Of particular concern with large ring laser gyros are the changes in Sagnac frequency due to changes in the relative phase and amplitude of the backscatter phasor. These can be driven by changes in geometry which move the standing wave pattern relative to the scattering centres. In principle it is possible to get some information on the backscatter phasor by monitoring the uncombined beams for *monobeam modulation* at the Sagnac frequency. Such modulation is readily observable in large ring lasers such as G-0 and PR-1 but not in ultra large ring lasers. Sagnac frequency correction using monobeam signals is an ongoing research topic.

2.3.4. Irreducible laser noise

The quantum limit[2] to the frequency uncertainty in a laser is a well studied problem in laser physics. The investigations began with Schawlow and Townes who presented what is now known as the Schawlow-Townes equation in their 1958 paper [22]. Since then a number of different authors have presented alternative analysis of the problem, often with very different approaches, yet yielding the same result.

While the approaches may vary, the fundamental assumption behind derivations of the quantum limit considers the spontaneous emission which inevitably accompanies laser gain in a laser operating in the steady state condition. This spontaneous emission causes a perturbation to the phase and polarisation of the beam.

[1] The first (and smallest, to the authors knowledge) ring laser gyro reported to unlock on earth rotation alone was the C-1 laser which enclosed an area of $0.75\,\mathrm{m}^2$ [21].

[2] Also known as the 'shot-noise limit'.

A review of the quantum limit in terms of the rotation rate measurement of ring laser gyros has been presented by Stedman [4], in which it is shown that the limit to the sensitivity of rotation rate measurement is

$$\Omega_q = \frac{cP}{4AQ}\sqrt{\frac{hf_0}{P_i t}} \tag{2.14}$$

where t is the measurement time and the other variables take their usual values[1]. Note that similar analysis has also been presented by other authors, in particular [23–26].

The quantum limit effectively imposes 'white' frequency noise on the rotation rate signal. As expected for this type of noise, the result can be improved by increasing the measurement time. The quantum limit can also be reduced by increasing the scale factor, although there may be practical limits to this (see Section 2.4.4). The main contributor to the quantum limit is the reflectivity of the mirrors, which in turn determines the cavity quality factor Q and allows the intra-cavity power to be increased. The reflectivity is determined by the quality of the surface preparation and deposition of the dielectic layers. There exists a theoretical lower limit to the losses in SiO_2 and TiO_2 multi-layer surfaces of the order of 10^{-9} [27].

In practise the noise in our large ring lasers is dominated by other factors which will be detailed in the following sections. To date we generally operate well above the quantum limit. However with other laser configurations operation near the quantum limit has been reported [26].

2.4. Performance of large ring lasers

When explaining large ring laser gyros to the lay person, one of the most difficult aspects to convey is their phenomenal sensitivity. A common factoid often given is that the angle of rotation resolvable over around a second of observation is equal to that subtended by a human hair[2] held at a distance of 500 km. While impressive to the lay person, these types of comparisons do not very well indicate what sets our systems apart from other types of gyros. The most important aspect of our

[1] See list of nomenclature, page 283.

[2] This factoid surreptitiously incorporates the large uncertainty in an estimate of this nature, because the width of a hair can vary by up to an order of magnitude.

measurements is their long term stability, in particular the length of time over which we can average while continuing to reduce the measurement error.

2.4.1. Characterisation by time series analysis

The easiest characterisation of the performance of a ring laser gyro is made simply by looking at a time series of the Sagnac frequency f_s, typically after taking a moving average. This allows any long term drifts to be identified.

Figure 2.6 shows an example plot of a long UG-3 data set illustrating a downward trend.

A constant downward drift in the Sagnac frequency is observed in both UG-2 and UG-3 data. This drift 'resets' after the gas is refilled, which indicates that it is a result of the accumulation of contaminants from outgassing in the gas mixture. The mechanism for this drift is explored in Section 3.6.2.

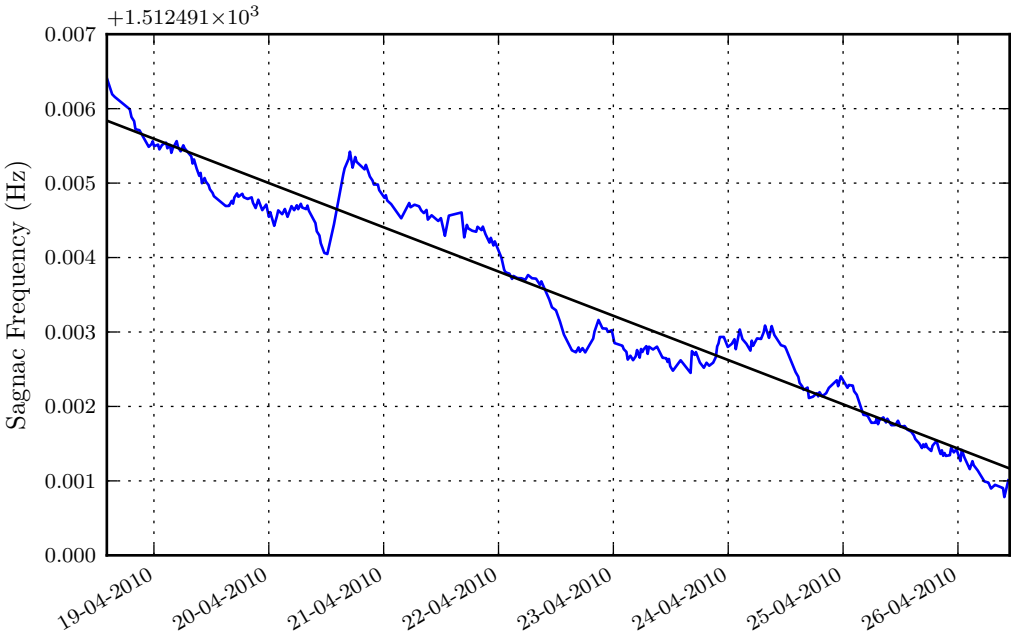


Figure 2.6. Illustration of downward drift of UG-3 raw Sagnac frequency with a linear fit. The ∼0.7 mHz per day drift is easily visible over a week of data. Dates represent UTC.

In UG-2 the partial pressure of hydrogen reaches approximately 0.02 mbar after ap-

proximately 55 days. At this point the available gain has decreased to the point where the laser can no longer be operated and it must be refilled. The rate of pressure rise in UG-3 is comparable and ultimately limits observation time in these instruments.

2.4.2. Characterisation of stability by Allan variance

In 1966 Allan [28] described a two-sample variance technique intended for quantifying the stability of clocks and oscillators. This is now know as the Allan Variance[1] and has been used extensively for characterising the stability of ring laser gyros. The basic procedure is to take blocks of n successive measurements of a quantity y and find the Allan variance

$$\sigma_A^2(n) = \frac{1}{2(n-1)} \sum_{p=1}^{n-1} (y_{p+1} - y_p)^2 \,. \tag{2.15}$$

The number of measurements in the block n is usually scaled in terms of measurement time in seconds. The Allan variance is calculated for successively larger measurement times and presented on a log-log plot with σ_A^2 on the vertical axis and measurement time on the horizontal axis. Multiple estimates made with different blocks of data of length n are normally averaged to obtain σ_A^2. When applied to large ring laser gyros, it is important to note that the measurement conditions can change from day to day. Ideally, Allan variance curves should be calculated from multiple independent runs and the results averaged for each measurement block length. This also allows error bars to be assigned from the standard deviation of the measurements for each run[2].

The Allan variance calculation can be implemented quite succinctly with any numerical programming language; good use of array reshaping routines can eliminate direct iteration and speed up the calculation. An example Python implementation is shown in Listing 2.1.

The Allan variance can be used to assess the effects of noise processes on the stability of the measurements. It gives no useful information about systematic errors or long term drifts. Different slopes on the Allan variance curve are characteristic of different classes of noise process. An overview of these classes and the corresponding Allan variance curves can be found in standard reviews such as Allan's 1987 review of the

[1] Standard abbreviation AVAR.

[2] This is done for the G-0 and UG-3 Allan variances presented in Figure 2.7.

technique [29]. Another useful overview, written in the context of ring laser gyro performance analysis was made by Ng [30].

Sometimes the alternative formulation, *Allan deviation*[1] is used, this is simply obtained by taking the square root of the Allan variance. The vertical axis of the Allan variance plots is sometimes given simply in terms of Sagnac frequency squared, or scaled appropriately to be in terms of rotational velocity squared or (as used in this thesis), fractional sidereal rotation rate squared.

Figure 2.7 shows Allan variance plots characterising the UG, G-0 and G ring lasers. The overall trend is an initial slope of around -2. This is indicative of white noise in the phase of the signal which is predominantly injected by microseismic motion. These motions also explain the excess of noise between 1 s and 5 s. The Allan variance shows a 'turn around' at around 500 s measurement time for the UG lasers, somewhat earlier for G-0 and much later for G. Beyond this turn around point, the measurement is no longer improved by further averaging.

After the turn around, the Allan variance has a slope of around $+1$ for the UG lasers, indicative of a 'random walk' noise process. The cause of this is probably very long term drifts associated with geometric scale factor changes or null shifts ultimately caused by gas contamination.

The G-0 laser shows a steeper slope beyond the turn around, tending to around $+2$. This is classified as 'frequency ramp' noise. The physical origin of this noise source is unknown.

Overall, we observe a very similar shape to the Allan variance curves of UG-2 and UG-3. This is expected because they are subject to the same noise sources. However, UG-2 is more sensitive to the noise than UG-3, and overall the performance is worse. The effects we believe to be responsible for this reduction in performance can be briefly summarised:

- The approximately square geometry of UG-1/UG-3 reduces changes in the scale factor caused by area changes associated with horizontal movements of the beams on the mirrors.
- The laser spot subtends a much larger area on the UG-2 mirrors than on

[1] Standard abbreviation, ADEV.

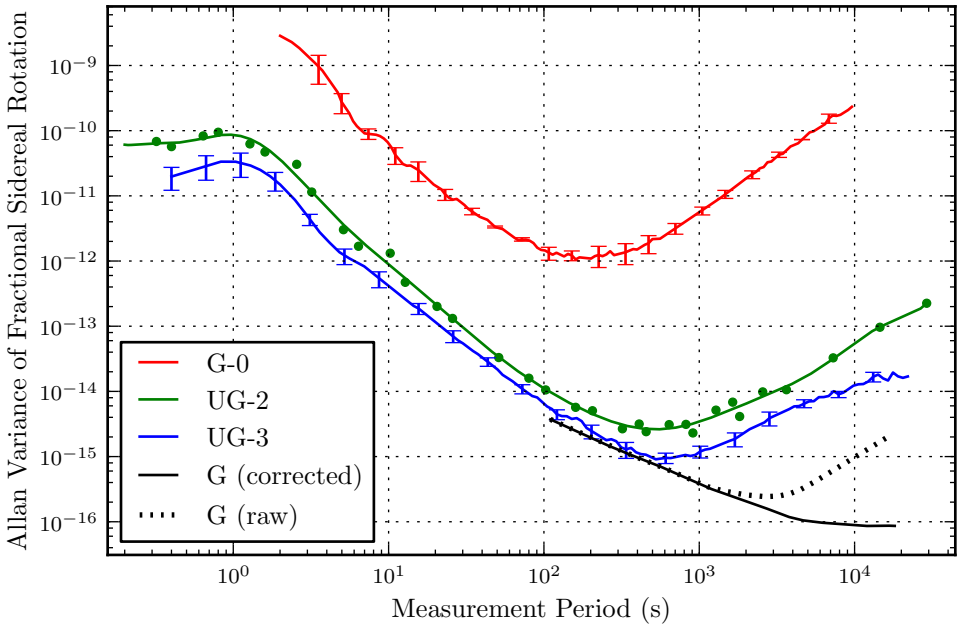


Figure 2.7. Allan variance comparison of UG-2, UG-3 and G-0 ring lasers.

```
1  import numpy as np
2
3  def avar(signal, n):
4      # Truncate the signal so it can be shaped to fit into blocks of length n
5      length = np.shape(signal)[0]
6      if (length % n) > 0: signal = signal[0: -1*(length % n)]
7      newshape = (np.shape(signal)[0]/n,n)
8      signal = np.reshape( signal, newshape )
9
10     # Find Allan variances for each block
11     means = np.mean(signal, axis=1)
12     avars = np.diff( means )**2
13
14     # Return the average Allan variance for all the blocks
15     return 0.5*np.mean(avars)
```

Listing 2.1 Example Python implementation of the Allan variance.

UG-1/UG-3. Additionally, the spot moves across the mirror surface at a greater rate with UG-2 given the same angular rate of change in alignment. The causes greater variations in the backscatter from the mirrors.

- UG-2 had higher losses than UG-1/UG-3. As a result it was necessary to increase the excitation power and consequently the temperature of the plasma increased. This increases any temperature related null shift effects such as those resulting from gas flows. It is possible that this is responsible for the greater uncertainty in the Allan variance of UG-2 for long measurement periods.
- Because of the higher losses, it was necessary to use a relatively narrow gain tube with UG-2. This increases the fraction of the beam which is obstructed by the edges of the plasma tube, and consequently increases the backscatter.

The performance of G initially tracks that of UG-3, with a small improvement. This is most likley because both instruments are limited by microseismic activity in this range and the microseismic activity is weaker at the further inland (Bavaria, Germany) site compared to UG-3. Unlike UG-3, the G Allan variance continues to improve well beyond 1000 s measurement periods. This is afforded by the much improved geometric stability of the G ring laser due to its monolithic construction. Additionally, the G ring is equipped with hydrogen 'getters' which improve gas life and eliminate long term drifts associated with temperature dependent null shifts.

Figure 2.7 also shows the Allan variance for a data set from the G ring to which a suite of corrections have been applied[1]. These include corrections for the Oppolzer terms (which account for diurnal polar motion), length of day corrections, solid earth tidal tilts and tilts induced by ocean tides. With these corrections applied it can be seen that the performance continues to improve with no clearly defined 'turn-around' for the length of the data set.

2.4.3. Characterisation of noise sources by autocovariance

While the Allan variance gives some indication of the frequency bands in which noise is present, it is not particularly well suited to this.

The technique favoured here is the autocovariance. This works by taking a copy of

[1] The author acknowledges Dr. André Gebauer for providing the time series data from from which these Allan variance curves were computed.

a signal $y(t)$ and time shifting it by some period τ. The signals are then multiplied together and averaged for different time shifts, i.e.

$$K(\tau) = \frac{\Delta t}{t_{\max}} \sum_{t=0}^{t=t_{\max}-\tau} \left(y(t) - \overline{y(t)}\right) \left(y(t+\tau) - \overline{y(t+\tau)}\right). \tag{2.16}$$

$K(\tau)$ can then be plotted to form an *autocovariance* function. This shows positive peaks corresponding to the period of any repeating element in the signal, sinusoidal or otherwise. The autocovariance is in principle an alternative display of the information contained in the associated power spectral density. The Wiener-Khinchin theorem shows how the autocovariance can be obtained from a Fourier transform of the power spectral density.

One advantage of the autocovariance is that it can be easily calculated at higher resolution than implied by the time steps in the individual measurements. This can be done by generating a cubic spline (or other suitable interpolator) from the data set and evaluating this when computing the autocovariance.

Often we would like to determine the precise period corresponding to a peak. This can be achieved by applying a localised optimisation technique, beginning at the approximate value of peak. This is a very efficient way of obtaining the precise period of a repeating signal. Note that Fourier based techniques would require re-sampling the whole interpolated signal at high resolution. Computationally this can be very memory intensive.

Figure 2.8 shows an autocovariance function computed from a week of UG-3 data. Repeating signals are visible, in particular with periods of 4.0 s (primary), 9.1 s and 13 s. These signals are an effect of the *microseismic background* and are the fastest varying rotation signals we observe. The main component of the microseismic background originates as pressure oscillations on the sea floor. These pressure variations are the result of interactions between oppositely travelling ocean waves with the same frequency [31, 32].

The exact period and amplitude of the microseismic motion varies considerably from day to day, but always has frequency of between 250 mHz and 450 mHz. Microseismic signals are further discussed in Section 5.1.

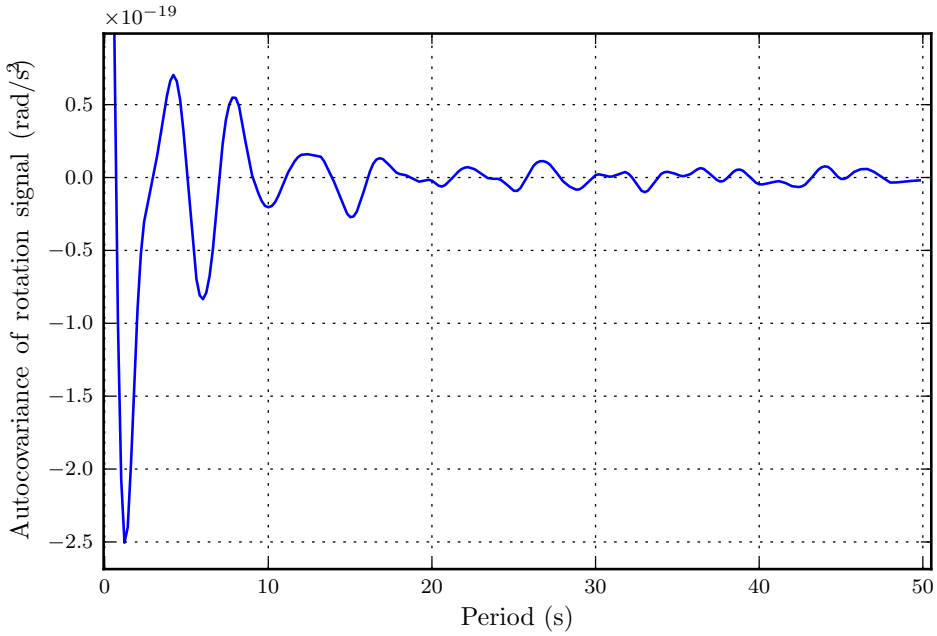


Figure 2.8. A typical autocovariance periodogram computed from a week of UG3 data, showing microseismic components for different periods (lag values in the autocovariance). Tends to $6 \times 10^{-19}\,\mathrm{rad}^2\,\mathrm{s}^{-2}$ at zero lag.

2.4.4. Practical limits to the scale factor

As we have seen in the earlier sections, an increase in scale factor of UG-2 compared to UG-1 was not accompanied by a corresponding increase in performance.

Generally speaking, the largest factor determining performance of a ring laser gyro is the mirror loss. Observed mirror losses are typically greater than the manufacturers specifications; the additional losses can only be accounted for by scattering. The probability of part of a beam falling on a scattering particle or defect increases as the area of the spot increases.

Comparing the spot sizes and losses[1] for recently constructed ring lasers we see a tendency for the average mirror loss to increase as the beam spot size at the mirrors increases and becomes subject to greater overall variability across the mirror surface. This is illustrated in Figure 2.9, here the lowest losses achieved have been plotted

[1] Found from ringdown times.

against spot size[1] for various lasers.

Our observations suggest that best performance should be obtained by minimising the spot areas, within the limits imposed by the necessity to have a stable cavity and to allow the beams to fit through the gain tube. However, even if this is done, we expect that for cavities above a certain 'optimum perimeter', the cavity Q will not increase further (and may actually decrease) as the cavity dimensions are increased due to the increasing spot area. Based on our experience with UG-2 and UG-3, we expect the optimal spot size given current mirror technology to be not much larger than 5 mm in $1/e^2$ (intensity) diameter.

The mirror manufacturers[2] specify a *clear aperture* diameter of typically 3 mm, this is the point in the centre of the surface which is guaranteed to be within specification. Ideally we would like to maintain the beams inside this area. It should be possible to develop an automatic alignment system to maintain the laser alignment to achieve this. Such a system has sometimes been referred to as a *virtual monolith*. It is likely that use of such a system could effectively increase the optimum spot area.

2.4.5. Operating wavelength

When discussing practical limits to the scale factor, it is interesting to consider the possibility of operation at shorter wavelengths.

Let us consider operation of a helium-neon ring laser on the 543.3 nm transition. This directly increases the scale factor by 14%. However, the laser spots on the mirrors are smaller. This allows the perimeter to be extended by 10.2% while maintaining the same spot size as we would have at 632.8 nm. Cumulatively, this allows an overall improvement of 20.6% in scale factor.

Operation on this wavelength represents a minimal departure from existing experience. Probably the biggest difference is the reduction in gain available at this transition by approximately a factor 30 [33]. In addition, assuming the spot sizes remain the same and that the losses are dominated by scattering and mirror defects, overall losses would be expected to increase for shorter wavelengths and scale with

[1] From the $1/e^2$ intensity contour. Note that for configurations where different spot areas have been used on the same laser, the geometric mean of the areas has been used.

[2] Research Electro-Optics, Inc.

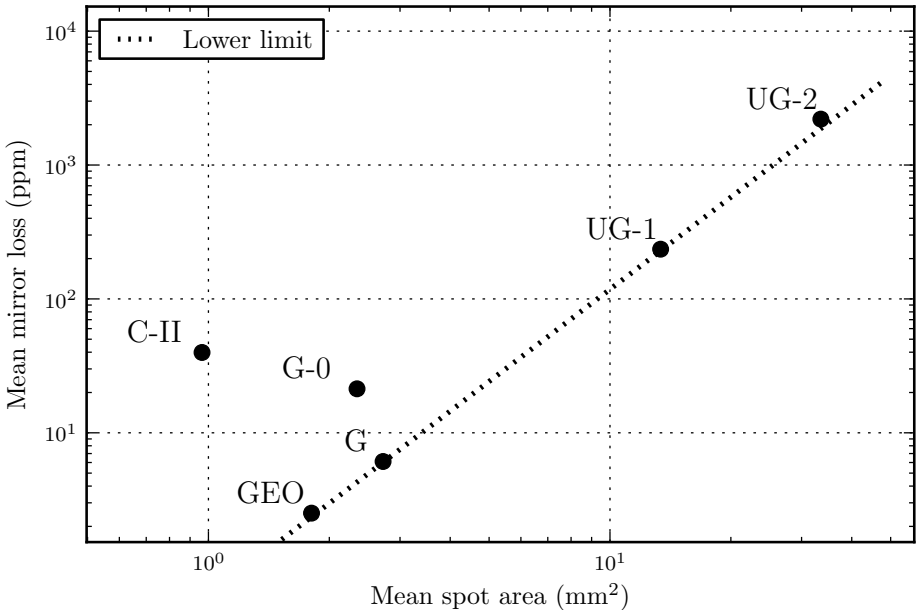


Figure 2.9. Illustration of increasing losses with spot area for a number of lasers.

λ^{-2} (see Section 2.3.3).

If operation at 543.3 nm is to be considered, it is recommended that the additional gain be supplied by lengthening the gain tube or installing multiple gain tubes. While it may be possible to operate a conventional gain tube at higher power, this may introduce undesirable temperature dependent null shifts.

Because of the cumulative improvements that can be achieved by reducing the wavelength, it is well worth considering operation on less familiar wavelengths, especially if instruments with perimeters exceeding 100 m are to be considered. A possible candidate might be the helium-cadmium system operating at 325 nm or 442 nm. In principle, a ring laser gyro with perimeter of the order of 250 m operating at 325 nm could be built with a reasonable 17% increase in spot size over UG-3. Such an instrument would have around a 600% improvement in optimal scale factor.

2.5. Alternative methods of operation

A number of alternative methods for operating ring laser gyros have been tried. These have mostly been applied to the small ring lasers which are designed to be

used as part of inertial navigation systems. Unlike our large ring laser gyros, these instruments can not unlock on earth rotation alone. They require non-reciprocal elements or moving parts to avoid lock-in.

Because of the commercial applications of small ring laser gyros, development of techniques to avoid lock-in has mostly been done behind the closed doors of companies such as Honeywell, Raytheon and Sperry [4, 34–36]. Nevertheless, the basic techniques are well known. Some of these may have applications for large ring lasers so a brief review is warranted.

2.5.1. Mechanical dither

Perhaps the simplest approach is to artificially move the region of measurement outside the lock in band. This can be done by applying an oscillating angular movement or *dither* (typically with amplitude of a few hundred arc seconds and frequency a few hundred cycles per second [25]) via a magnetic actuator [37], piezo drive [38, 39] (as part of a resonant mechanical system), or by tilting the mirrors [40]. This introduces an alternating null-shift and as a result the gyro spends only a small fraction of time in the lock-in band and the average rotation rate is to first order unaffected by lock-in.

The main limitation to this approach is that the sensitivity depends on the accuracy to which the dither can be controlled. Control of the dither motion becomes increasingly difficult as the gyro becomes larger. Additionally, it destroys the intrinsic simplicity of the ring laser gyro which otherwise has no moving parts.

In a 1998 review of the preceding 40 years of inertial navigation technology, King [41] remarked:

> "I suspect that any advantages of the multi-oscillator technique are balanced by corresponding difficulties and, like the dynamically tuned gyro versus the floated gyro, the overall effect on the evolution of inertial navigation technology will not be too significant. Our own ring laser gyros continue to use mechanical dither, which is robust and reliable."

This shows that despite the disadvantages, dithered gyros are sufficient in most navigation applications.

2.5.2. Fresnel-Fizeau

The *Fresnel-Fizeau effect* is a non-reciprocal change in the refractive index of a moving dielectric medium. If a medium of length l is established with constant velocity v in a ring laser cavity, the induced frequency splitting f is given by

$$f = \frac{2(n^2 - 1)vl}{\lambda P} \tag{2.17}$$

where n is the refractive index of the medium [17]. This essentially introduces a null shift which can bias the gyro out of lock-in.

Both spinning glass wheels [42, 43] and spheres [44] have been used to demonstrate the effect. As with other mechanical techniques, maintaining the necessary precision in controlling the offset is the limiting factor. To the authors knowledge, this has never been applied beyond the 'demonstration of principle' stage.

2.5.3. Multi-oscillator

The *multi-oscillator* approach requires introducing a frequency or phase bias to the co-rotating beams dependent on the direction of travel. The approach is elegant in principle, but difficult to implement in practice.

The most common method (and to the authors knowledge, only one to be put into practice commercially) makes use of Faraday rotation. The basic principle is illustrated in Figure 2.10.

The first stage of the process requires introducing an optical element into the ring cavity which causes a different phase shift for left and right hand circularly polarised light, such as a quartz crystal. The second stage makes use of a material with high Verdet constant to which a magnetic field has been applied (Faraday rotator). Overall this introduces a different phase shift for the clockwise and counter-clockwise beams. Consequently the co-rotating beams have quite different optical frequency. They are each subject to the Sagnac splitting in the usual way.

Polarisation selecting optics can be applied to a combined output beam to obtain two beat frequency outputs. Subtracting these two cancels the applied bias and allows the Sagnac frequency to be recovered.

Because most mirrors will preferentially select (linear) s-polarisation, it is necessary

to apply $\frac{1}{4}\lambda$ waveplates to each end of the Faraday rotator so that only linearly polarised light is incident on the mirrors.

While this approach is elegant in principle, it is very complicated in practice and as such has numerous technical challenges. It requires a large number of optical elements to be placed in the laser cavity which increases the losses and has a detrimental effect on the quantum limited rotation rate. Most materials with a high enough Verdet constant sufficient for use as practical Faraday rotators have poor thermal properties. Nevertheless, Litton Industries (now owned by Northrop Grumman Corporation) have been successful in producing small ring laser gyros operating using this technique [14].

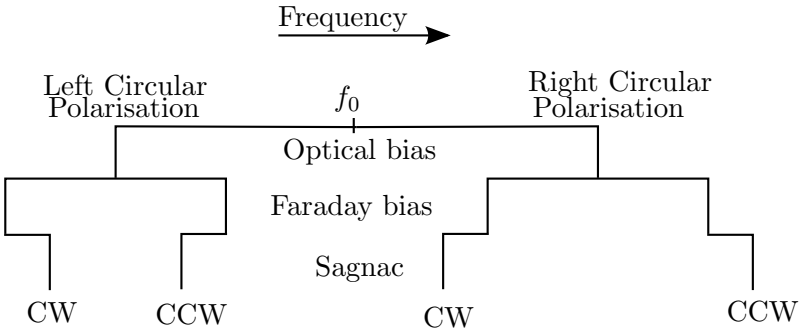


Figure 2.10. Frequency biases in the multi-oscillator ring laser gyro.

2.5.4. Transverse magneto-optic Kerr effect

The *transverse magneto-optic Kerr effect* is the phenomenon by which p-polarised light reflected from a material magnetised perpendicular to the plane of incidence encounters a small phase shift [45] dependent on the direction[1] of magnetisation and a corresponding change in the reflectivity of the surface. There is no change in the polarisation of the reflected beam. The effect is probably most familiar through its application in magneto-optical drives for data storage, where the change in reflectivity is exploited.

The effect can be understood through a classical model of light propagation through a material where the oscillating electric field displaces bound electrons which can be considered equivalent to a localised current flow. The presence of a magnetic field

[1] I.e. the sign of the magnetic field vector.

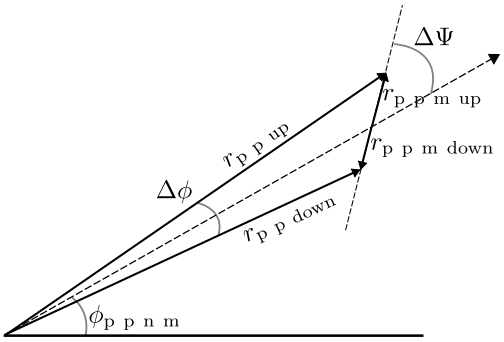


Figure 2.11. Illustration of incident and reflected electric field vectors from a magnetic mirror for both opposite magnetisations. Subscripts p p indicate p-polarisation, m indicates a magnetically derived component, n m indicates a non-magnetic component.

gives rise to an additional component of the associated electric field. Figure 2.11 illustrates how through addition of the non-magnetic (dashed) component with a vector resulting from the magnetisation, a phase shift $\Delta\phi$ is observed in the reflected beam.

An alternating electric field can be used to alternate the phase shift, effectively applying a dither in the same way as is done with a dithered mechanical gyro (see Section 2.5.1). This type of gyro is sometimes known as a 'magnetic mirror' gyro.

This can be implemented in practice by coating one (or more) of the mirrors with a magnetic material in which the direction of the magnetic dipoles is easily switched (a 'soft' magnetic material). Transparent bismuth-substituted yttrium-iron garnet is an example of such a material that has been used [46].

The main potential advantage of this method is that if a sufficiently large square-wave field is used to quickly drive the surface into magnetic saturation, then the magnitude of the phase shift is fundamentally fixed. There is no specific requirement for high stability of the field. Note that the requirement for high quality mirrors remains, this was shown by Wilkinson [25] who noted that in the presence of backscatter the null shifts do not integrate to zero.

This technique has been demonstrated by Andrews and others [46, 47] with some success. In experiments to date however, the main disadvantages of this method have been cited as the constraints on the exact reversibility of the bias and the low

levels of mirror loss and scatter required [48]. The performance of this type of gyro has not been shown to reach the level required for navigation. In 1996 Andrews *et al.* [48] noted that considerable improvements in magnetic mirror fabrication would be required to reach the values of stability and dynamic range already achieved by conventional gyros.

Nevertheless, this method is in principle quite compelling and may be worthy of a second look given recent advances in mirror fabrication technology. One can imagine an exotic mirror incorporating a suitable magnetic material and solid-state gain medium directly as part of the mirror structure to form a very elegant system. Layered gain structures in ring lasers have been investigated in a different though related context by Diels [49] and Scully [50]. In these examples the main topic being investigated was exploiting destructive interference in the layers to reduce spontaneous emission.

3. Gain in the Helium-Neon Laser

In this research, (and in almost all earlier large ring laser gyro research) the 632.8 nm transition of the helium-neon laser system has been used. As the stability of our instruments continues to improve the ability to accurately model the gain process and understand the subtleties particular to the helium-neon system becomes increasingly important.

The goal of this chapter is to describe the system and its subtleties, backed up by experimental measurements where necessary. The processes underlying the generation of the gain curve are of particular importance. Understanding these can give insight into factors limiting the performance of ring laser gyros, and allow corrections to be made. It could also allow better selection of operating parameters such as gas pressure and mixture; in the past these parameters have been selected only partly on a theoretical basis, and partly by rules of thumb.

3.1. The helium-neon laser system

Figure 3.1 shows the basic atomic transitions which give rise to the 632.8 nm emission. The pumping mechanism, (in our case, free electrons accelerated by a RF electric field) bring helium atoms to a long lived metastable energy level of 20.61 eV. The lifetime of this state is 2.04 ms, [51, 52] which is very much longer than all other relevant lifetimes. By good fortune there is a close coincidence in energy levels between the He 2^1S level and the Ne 5s levels.

With the addition of around 0.05 eV of thermal energy[1], the helium atom can undergo collisional energy transfer with a ground state neon atom, leaving the neon atom in one of the 5s states. In our case we are interested primarily in the upper 5s state

[1] Coincidently, the average translational kinetic energy at a typical operating temperature of 400 K is 0.052 eV.

(denoted 5s′ in Figure 3.1). This state is sufficiently long lived (≈55 ns) to allow a population inversion to be established. Laser amplification can then occur at 632.8 nm and 543.3 nm [53] wavelengths from the 5s′ state to the J=$\frac{1}{2}$ and J=$\frac{3}{2}$ 3p states respectively. Here, J is the total angular momentum quantum number. Laser gain at 543.3 nm is a factor 30 smaller [33] than at 632.8 nm. Laser gain can be increased by approximately 20% by using ^{3}He rather than ^{4}He [33], due to a slightly better match between the energy levels.

After stimulated (or spontaneous) emission occurs we note that the neon atom is still in a 3p state. Transitions from the 3p level directly to the ground state (lifetime ∼19 ns) are unlikely as the many transitions from the 3p state to lower energy excited states are each much more probable (lifetimes of 7 ns to 10 ns [54]). In practice, spontaneous emission to one of the many 3s states, followed by collisions with the walls of the gain tube (which happen after between 10 ns and 100 ns, depending on discharge pressure and distance to the walls) are mostly responsible for the final transition to the ground state.

It is the slowness of these lower transitions that fundamentally limits the maximum power of any realistic helium-neon laser. Where maximum power is desired, the gain tube can be made as narrow as possible and pressure made as low as possible to increase the frequency of wall collisions. The most powerful helium-neon laser ever made (to the author's knowledge) [55] was capable of 350 mW output and used two flattened gain tubes to maximise wall area. Note that in our systems (neon partial pressure 0.2 mbar), the mean free path for neon is typically around 10 mm, i.e. comparable to the tube diameter.

3.2. Helium-neon gas mixture

It has been found empirically [58] that the optimum *pd* (pressure diameter) product is a few Torr mm, and that the gain per unit length varies inversely with tube diameter. Gordon and Wright [59, 60] have quoted empirical results for the optimal *pd* product of 14.6 Torr mm and 6.6 Torr mm, the former using a RF discharge and the latter using a DC discharge.

The single pass gain of a helium-neon laser is typically between 2%/m and 10%/m [58]. Herziger *et al.* [61] given an expression based on empirical measurements for the

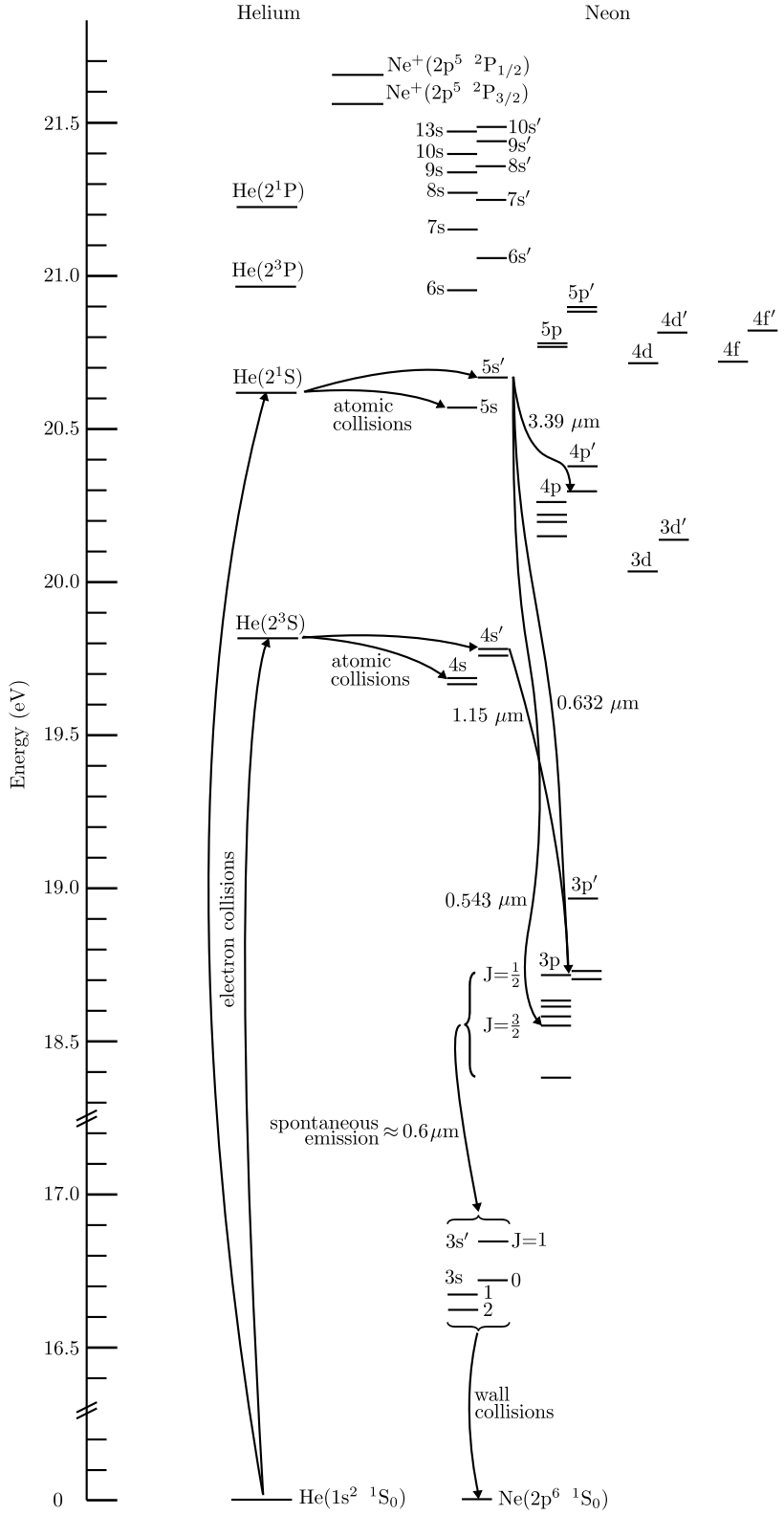


Figure 3.1. Transitions of the helium-neon system. Energy level information taken from Haberland [56] and NIST Atomic Spectra Database [57].

maximum single pass gain G_0 as a function of discharge tube length l and diameter d,

$$G_0(l,d) \approx \left[1 + \frac{1}{2}\left(\frac{d_0}{d}\right)^{1.4}\right]^{l/l_0}. \tag{3.1}$$

Here, l_0=1 m and d_0=1 mm. Note that they consider this expression valid for an 'optimal' gas mixture.

Among standard references [62–66], reported values of the optimal He:Ne gas mixture ratios vary between 5:1 and 20:1. These appear to all be based on empirical studies. In general it appears that there is relatively little dependence of the gain on the precise ratio. Also, it is likely that manufacturers of helium-neon lasers would overfill with helium slightly, anticipating the depletion of helium by diffusion out of the cavity over the lifetime of the laser.

Mielenz and Nefflen [67] have measured the optimal pressures for different RF excited He:Ne gas mixture ratios at different tube diameters. Re-analysing their experimental results allows us to give an expression for the optimum neon partial pressure as a function of total pressure p (in mbar) and tube diameter d (in mm),

$$p_{\mathrm{Ne}} = \frac{p}{6.21p + 2.09d - 17.95}. \tag{3.2}$$

Strictly this is valid only for their measurements with d between 3 mm and 5 mm, and pressures between 1.8 mbar and 3.3 mbar. Given the self consistency of their results however, it is reasonable to apply this to our systems where tube diameters of up to 8 mm and pressures up to 8 mbar are often used.

As an example, Equation 3.2 gives an optimal neon partial pressures of 0.168 mbar and 0.165 mbar for total pressures of 5 mbar and 8 mbar respectively in an 8 mm gain tube. The difference here is almost within the uncertainty to which the gas can be mixed in practice.

In our helium-neon ring lasers, we have historically used a neon partial pressure of 0.2 mbar, independent of total pressure and tube diameter.

3.3. The helium-neon gain curve

The gain curve is a function that relates the amplification to the optical frequency in a laser gain media. Being able to construct gain curves for different laser configurations

is very important for understanding the behaviour of modes in the laser.

Let us first consider the *small-signal gain curve*, where the lasing intensity is low and any saturation processes are negligible. The spectral line of the lasing transition is subject to two classes of broadening processes, *homogeneous broadening* and *inhomogeneous broadening*.

3.3.1. Homogeneous broadening

Homogeneous broadening processes are so called because they apply to all atoms in the gain medium in the same way. The most fundamental of the broadening homogeneous processes arises directly from the Heisenberg uncertainty principle, which requires the frequency f to have some uncertainty

$$\Delta(hf)\,\Delta t \geq \hbar/2 \tag{3.3}$$

given a finite mean lifetime Δt of the transition. This results in a Lorentzian broadening of the spectral line with width (full width at half maximum) given by the decay rate of the transition known as the *natural linewidth* (here donated l_{w0}). It is usually simply the Einstein A coefficient, although is can be increased if there are any non-radiative decay processes (see Siegman 3.2 [58]). For the 632.8 nm helium-neon laser line, the lower limit to l_{w0} is 0.54 MHz (from the NIST Atomic Spectra Database [57]).

The dominant cause of homogeneous broadening in a gas laser however is *pressure broadening*[1]. This is caused by collisions of atoms which interrupt the emission process and cause a change in the lifetime of the atomic transition. This occurs either through 'hard' collisions that change the state of the Ne atoms, or via 'soft' collisions that change the velocity so that the atoms are no longer in resonance with the laser beam.

Again, the effect on the spectral line is described by a Lorentzian function. This arises from the atomic susceptibility function (see Siegman [58] 2.54), from which we

[1] Also known as collisional broadening.

consider only the normalised real part[1]. We denote the Lorentzian gain curve

$$L(f) = \frac{1}{1 + \left(\frac{2(f-f_0)}{l_w}\right)^2}. \tag{3.4}$$

Here, f_0 is the optical frequency and depends on the transition and isotope.

The Lorentzian line with l_w (full width at half maximum) depends on the mean time between collisions. It is usually found as a function of pressure from experimentally derived results.

Measurements from which the Lorentzian line width l_w can be found as a function of pressure have been reported by at least three groups. The most commonly used in ring laser research to date was presented by Smith *et al.* [68, 69],

$$l_w = 2\,(8.5 + 45p)\,. \tag{3.5}$$

Here l_w is in MHz and a pressure p is in mbar.

From the work of Tuchin [70], an expression can be found which gives l_w as a function of the partial pressures of helium p_{He} and neon,

$$l_w = 2\,((44.25 \pm 3)p_{\mathrm{He}} + (21.75 \pm 5)p_{\mathrm{Ne}} + (10 \pm 4)) \tag{3.6}$$

which is valid for gas pressures between 0.18 mbar and 3.6 mbar.

Knuston and Bennett [71] have presented experimental results which give l_w for pressures between 0.5 mbar and 4 mbar with He:Ne mixture ratios of 1:1 and 20:1. Crucially, these are given in terms of density[2]. After digitising their raw data and extrapolating linearly between the mixture ratios, we find an expression for l_w in terms of temperature T and fraction of helium H,

$$l_w = (0.23 \pm 0.02)\rho H + (35.5 \pm 0.8)\rho + l_{w0} \tag{3.7}$$

where ρ is 1×10^{-16} times the number density in atoms/cm, or

$$\rho = \frac{pN_A}{RT\,10^{20}} \tag{3.8}$$

[1] We are not concerned with dispersion in this analysis, but it comes from the imaginary part of the susceptibility function in the same manner.

[2] They report the measurements being made at a constant temperature of 295 K, although it is not clear how, or even if this was measured experimentally.

with pressure in mbar. From this model we can see that at typical operating conditions (6 mbar total pressure, 19:1 mixture), l_w can vary by 144 ± 30 MHz or 28% with a temperature shift of 100 K.

We note that while Kunston and Bennett's parameterisation is an improvement on other work, it is not strictly correct. The physics of the collisional broadening process suggests that the broadening should be proportional to the collision rate, which in an ideal gas is proportional to (number density)$\times\sqrt{T}$ or to $pT^{-\frac{1}{2}}$. Unfortunately no authors appear to have expressed the collisional broadening using this parameterisation.

When compared (see Figure 3.2), these models give results within their respective uncertainties. The difference in pressure broadening over the range of gas mixture ratios used in practice is comparatively small relative to the uncertainties of the models. In general the model based on Kunston and Bennett's data is recommended for future work because of the dependence on temperature and the smaller uncertainties than the model from Tuchin's data.

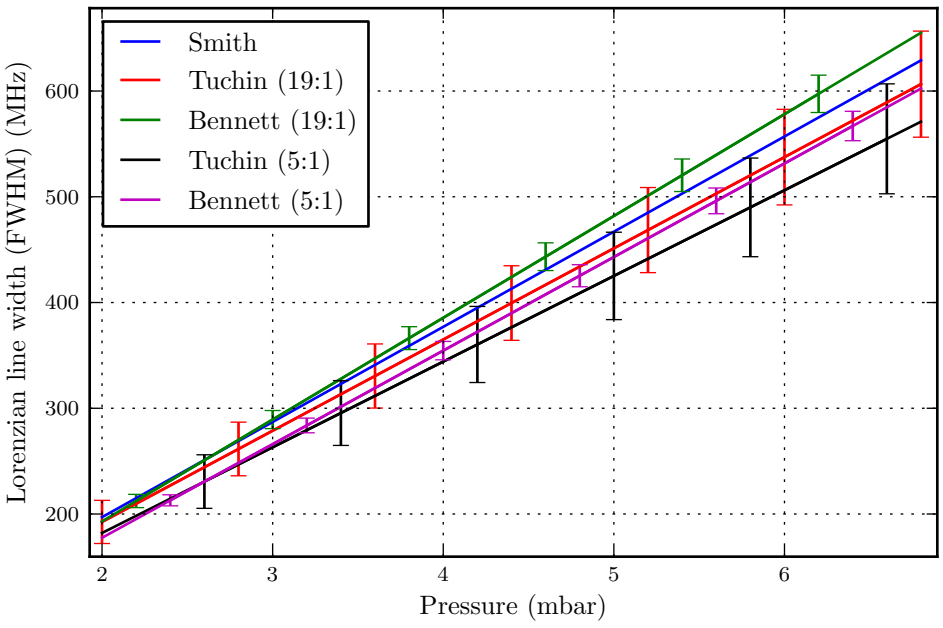


Figure 3.2. Comparison of different models for pressure broadening in helium-neon gas mixtures for different He:Ne ratios. Temperature was set to 300 K.

3.3.2. Inhomogeneous broadening

Inhomogeneous broadening processes are so called because they apply differently to different atoms in the gain medium. Doppler broadening is a standard example of such a process, and is the only inhomogeneous broadening process important to helium-neon gas lasers.

Doppler broadening arises because of the Maxwellian distribution of atomic velocities in the gain medium. This causes the transition to be broadened following a Gaussian line shape,

$$G(f) = \exp\left(-4\log(2)\left(\frac{f - f_0}{g_w}\right)^2\right). \tag{3.9}$$

Here the Gaussian line width[1] g_w caused by Doppler broadening is

$$g_w = f_0\sqrt{\frac{8\log(2)k_bT}{mc^2}} \tag{3.10}$$

where T is the effective gas temperature[2], m is the isotope mass and the other variables take their usual values[3]. As an example, at 400 K, a ^{20}Ne only mixture will have $g_w = 1517$ MHz.

3.3.3. Saturation and hole burning

In the analysis thus far we have considered the broadening mechanisms in the absence of an actual laser beam. In the presence of a laser beam however, *hole-burning* occurs. The mechanism is so named because it causes a localised depletion in the population inversion, giving a region of reduced gain or a 'hole' which is said to be 'burned' at the operating wavelength with 'depth' dependent upon the intensity of the beam.

This process was first described by Bennett [72]. Broadly, it can be described by a systematic removal of atoms falling within the Lorentzian spectral 'packet' centred on the frequency of the lasing mode. This occurs because the mode has caused these atoms to undergo stimulated emission. The effect becomes more pronounced as the intensity of the lasing mode is increased, or if the effective recovery time of the transition is increased.

[1] Also known as the inhomogeneous broadening width.

[2] The true 'effective temperature' depends on the temperature profile across the tube, weighted by the beam intensity profile. In most calculations this is not dealt with explicitly.

[3] See list of nomenclature, page 283.

Following the treatment of Siegman [58], we modify the inhomogeneous gain distribution by introducing a saturation function

$$S(f) = \frac{1}{1 + F(f)} \tag{3.11}$$

where

$$F(f) = \frac{I}{I_{\text{sat}}} \frac{1}{1 + \left(\frac{2(f - f_0)}{l_w}\right)^2} \tag{3.12}$$

which is essentially an inverted Lorentzian. Here I is the intra-cavity intensity and I_{sat} is the *saturation intensity*.

In a ring laser gyro, we have beams going in both directions. Each beam burns its own hole. The beams have equal (in magnitude) but opposite velocity, and therefore interact with populations of atoms on either side of the velocity distribution (unsaturated, Doppler broadened gain curve). Siegman [58] has considered the case of multiple beams in the analysis of the *Lamb dip*[1]. In this analysis, the saturation function is simply formed by adding the $F(f)$ (Equation 3.12) terms for each operating mode to form the overall saturation function.

Using Siegman's method of forming the $F(f)$ function also allows for situations where there is a power imbalance in the beams. This situation is readily observed in smaller ring laser gyros (C-II, G) and the mechanism for the imbalance remains unknown. In the larger instruments (UG-2, UG-3) however, the co-rotating beam powers have always been found to be equal within measurement errors.

In order to calculate the magnitude of the hole burning, it is necessary to at least know the ratio of the intra-cavity intensity I to the saturation intensity I_{sat}, though ideally we would like to know both parameters independently. The intra-cavity intensity can be found easily given a calibrated power meter, mirror transmission and beam geometry. The saturation intensity however is more difficult to obtain.

By definition, the saturation intensity is the intensity at which the gain drops to exactly half of the small-signal gain. Analytically the saturation intensity is simply

$$I_{\text{sat}} = \frac{hf}{\sigma \tau_{\text{eff}}} \tag{3.13}$$

[1]The Lamb dip is related to hole burning, but the 'dip' is in the output intensity of a laser being tuned through line centre when under constant excitation and occurs due to interacting 'holes'.

[58] where σ is the (energy dependent) transition cross section and τ_{eff} is the effective recovery time. In practice both of these parameters depend on temperature, pressure, gain tube diameter and gas composition in an involved way. It is better to use direct experimental measurements of the saturation intensity than to try to calculate it from the basic definition.

Smith [68, 73] measured the saturation intensity in DC excited tubes of varying diameter and pressure. This was done by monitoring the Lamb dip in the output power of a scanning helium-neon laser cavity. All tubes used a 7:1 mixture of He^3 and Ne^{20}. From a polynomial fit to their measurements we obtain

$$I_{\text{sat}} = 0.6509p^2 + 3.3604p + 1.3659 \tag{3.14}$$

in $\text{W}\,\text{cm}^{-2}$, for example $7\,\text{W}\,\text{cm}^{-2}$ at 1.33 mbar. This expression is purely phenomenological and valid up to 5.3 mbar ($37.5\,\text{W}\,\text{cm}^{-2}$).

Patel *et al.* [74] has also measured saturation intensity in a helium-neon laser. This was done by measuring the laser power deflected from an intra-cavity rotatable reflector. They report a much larger[1] $35.6 \pm 3.5\,\text{W}\,\text{cm}^{-2}$ in a 3.5 mm DC excited tube at 1.33 mbar. Unfortunately the pressure dependence and isotope mix is not given.

Note that in these measurements, the authors have expressed the intensities in terms of the slightly unusual $1/e$ intensity radius. Also note that the authors essentially assumed a uniform beam intensity profile. In a ring laser we usually have a TEM_{00} (Gaussian) intensity profile and as a result the gain medium saturates more readily at the locations corresponding to the more intense central parts of the beam. Siegman (8.3) [58] has considered this case and shown that introduction of a factor 2 to the beam area A_{eff}, i.e.

$$A_{\text{eff}} = \pi w^2 \tag{3.15}$$

where w is the $1/e^2$ intensity radius gives a better representation of the saturation intensity. The geometry of our lasers usually leads to significant astigmatism and we use

$$A_{\text{eff}} = \pi w_1 w_2 \tag{3.16}$$

with w_1 and w_2 the in-plane and orthogonal radii.

[1] Confusingly, they refer to this as 'slightly larger' than the measurements of Smith [73, 75].

Isotope	^{20}Ne	^{21}Ne	^{22}Ne
Mass (AMU)	19.992 440 176	20.993 846 74	21.991 385 50
Natural abundance	0.8949 ± 0.01	$0.0027 \pm .0001$	0.0958 ± 0.004
632.8 nm transition (MHz)	$473\,612\,251 \pm 10$		$473\,613\,149 \pm 10$
543 nm transition (MHz)	$551\,579\,659 \pm 50$		$551\,580\,659 \pm 50$

Table 3.1. Lasing transition frequencies [57, 76, 77], masses and abundances [78].

3.3.4. The Voigt gain curve

To calculate the overall gain at any particular frequency we must multiply the homogeneous response produced by each atom that has resonant frequency f_a by the fraction of atoms which share the resonant frequency due to inhomogeneous broadening. The easiest way to do this is to simply de-tune the Lorentzian gain curve so that it peaks at the frequency we want to find the overall gain at, then integrate the convolution of this with the inhomogeneous gain curve, i.e.

$$g(f) = \int_{-\infty}^{\infty} G(f_a) L(f_a, f)\, \mathrm{d}f_a, \tag{3.17}$$

which must be evaluated numerically. This results in a *Voigt* curve. Note that in most cases we will have multiplied the inhomogeneous broadening curve with the saturation curve $S(f_a)$. The resulting gain curve is not strictly a Voigt curve, but is often referred to as such.

The final gain curve for a mixture of neon isotopes can then be found by summing curves normalised for the relative neon isotope abundances.

The transition frequencies of the ^{20}Ne and ^{22}Ne isotopes are shown in Table 3.1. Note that the transition frequency for ^{21}Ne is unknown, but it can be ignored in most cases due to the low abundance. The exact values for the 632.8 nm transition were found from positions relative to known iodine absorption lines [76] which are known to high accuracy as they are frequently used length standards. The 543 nm transition isotope separations are less accurately known and were found from Gerstenberger *et al.* [77] with the ^{20}Ne wavelength found from the NIST atomic spectral line database [57].

3.3.5. Calculation of gain curves

The calculation of gain curves for different mode configurations and gas parameters is a common task in this research. We often want to compute gain curves at high resolution, or apply an optimisation routine to find the gas parameters corresponding to a particular behaviour. Because evaluation of each point on the gain curve requires numerical integration, such tasks can be relatively computationally intensive.

To simplify the calculation of gain curves, a Python module, *gaincurve* and associated module *physconst* has been written which incorporates the equations and various numerical models presented in the previous sections. Full documentation and usage information is available in Section A.1.1. Listing 3.1 show an example usage for generating a family of gain curves (Figure 3.3) showing the effect of increasing intensity on the hole-burning in a gain curve of a natural neon isotope mixture at 3 mbar.

The specific implementation has gone through several iterations and for this reason a brief discussion is warranted here.

The main challenges have largely involved finding an implementation that can transparently combine different techniques which give high accuracy or fast performance, each of which are required in different use-cases. Also it is important to keep the underlying physics clear to the user so that results from any part of the computation can be isolated and easily checked.

The basic idea is to define a set of operators which can be used to assemble a gain curve by chaining together objects representing the more elementary functions (e.g. Gaussian and Lorenzian curves). The advantage of this is that evaluation can be left until after the whole curve has been defined.

All of the objects used for building up the gain curve are simultaneously compatible with both 'continuous' and 'discrete' evaluation. For example, the 'continuous' evaluation option can be used when evaluating an integral. In this case an integration routine is given a function which it then evaluates at various discrete values in order to converge to a solution. This approach essentially gives a solution free from discretisation artefacts, although it tends to be computationally expensive. The same integral using discrete evaluation can be an order of magnitude faster and can be done in parallel. In this case, a range of equally spaced values would be sampled and

the integral found from these. In our implementation this is done using a Romberg integration[1] on $2^k + 1$ evenly spaced, pre-sampled points. Figure 3.4 shows an example where the two techniques have been used to evaluate a Voigt gain curve. Note that for good results k between 12 and 16 would normally be used. This limits computation times for a typical gain curve to less than a second on a standard PC.

3.4. Longitudinal mode configurations

Given typical gas parameters, the helium-neon gain curve is several GHz wide. Even in a small laser, many mode spacings can fit under this gain curve. A ring laser gyro is (in the simplest picture) unable to produce a useful beat signal with multiple modes running, and for this reason the common objection when larger helium-neon ring laser gyros were proposed was that it would be increasingly difficult to achieve single mode operation as the mode spacing decreased. However in practice this was relatively straightforwardly achieved. It was made possible by:

1. Operating just above the lasing threshold (*gain starvation* approach [4]) so that the modes neighbouring the mode nearer line centre were just below threshold.
2. Increasing gas pressure so that the spectral hole 'burned' by the operating mode was wide enough to suppress neighbouring modes and thus allow single mode operation over a wide de-tuning range.

It is not necessarily true that the operating mode will take the frequency of the nearest possible multiple of the mode spacing to the centre (maximum) of the gain curve. Because the gain curve is sufficiently flat over the range of a few mode spacings, hole-burning caused by the operating mode can suppress the gain of potential neighbouring modes and a stable configuration can result. On the UG-3 laser for example, stable configurations have been observed up to around 20 MHz from the centre of the gain curve. This phenomenon is particularly useful in a heterolithic ring laser because it affords a significant potential de-tuning range as the perimeter of the laser changes.

[1] Romberg integration uses the trapezoid rule at step-sizes related by a power of two and then performs Richardson extrapolation on these estimates to approximate the integral with a high degree of accuracy. See Numerical Recipes [79].

```
import gaincurve as gc
import matplotlib.pyplot as plt

plt.hold(True) # Put all plots on the same axis

# Make the gas mixture
gas = gc.LaserGas(400.0, "5s'->3p J=1/2")
gas.add_species('Ne 20', 0.2*0.9, True)
gas.add_species('Ne 22', 0.2*0.1, True)
gas.add_species('He 4', 5.0, False)

f0 = gas.unsat_max
fl = f0 + 15.2 # Some arbitrary lasing frequency

for isat in [0, 0.1, 0.2, 0.3, 0.4, 0.5]:
    # Generate the Lorentzian to use for integrating.
    lor = gc.Lorentzian(gas, 0.0)

    # Make the Gaussian
    gauss20 = gc.Gaussian(gas, 'Ne 20')
    gauss22 = gc.Gaussian(gas, 'Ne 22')

    # Generate the holes and apply to the Gaussian curves
    sat = gc.Saturation(gas, (f0+(fl-f0), f0-(fl-f0) ), (isat, isat), 1.0)
    gauss_hole20 = gc.Convolve( (gauss20, sat) )
    gauss_hole22 = gc.Convolve( (gauss22, sat) )

    # Now use both of these to make a Voigt profile
    vgt20 = gc.Voigt( gauss_hole20, lor)
    vgt22 = gc.Voigt( gauss_hole22, lor)

    # Scale the profiles and add together
    scaled_vgt20 = gc.Scalecurve(vgt20, gas.get_lasing_ratio('Ne 20'))
    scaled_vgt22 = gc.Scalecurve(vgt22, gas.get_lasing_ratio('Ne 22'))
    composite = gc.Sumcurve( (scaled_vgt22, scaled_vgt20) )

    # Evaluate the composite curve at discrete (50 MHz) points and plot
    composite.evaluate(7000.0, f0, 50.0)
    plt.plot(composite.x, composite.y, 'k-')

plt.xlabel('Frequency (MHz)')
plt.ylabel('Gain (arbitrary)')
plt.grid(True)
plt.show()
```

Listing 3.1 Example code snippet demonstrating the use of the *gaincurve* module. Output is shown in Figure 3.3. Note the convenience method `LaserGas.make_composite()` could be used to write this mode compactly, however it is written out in full here for illustrative purposes.

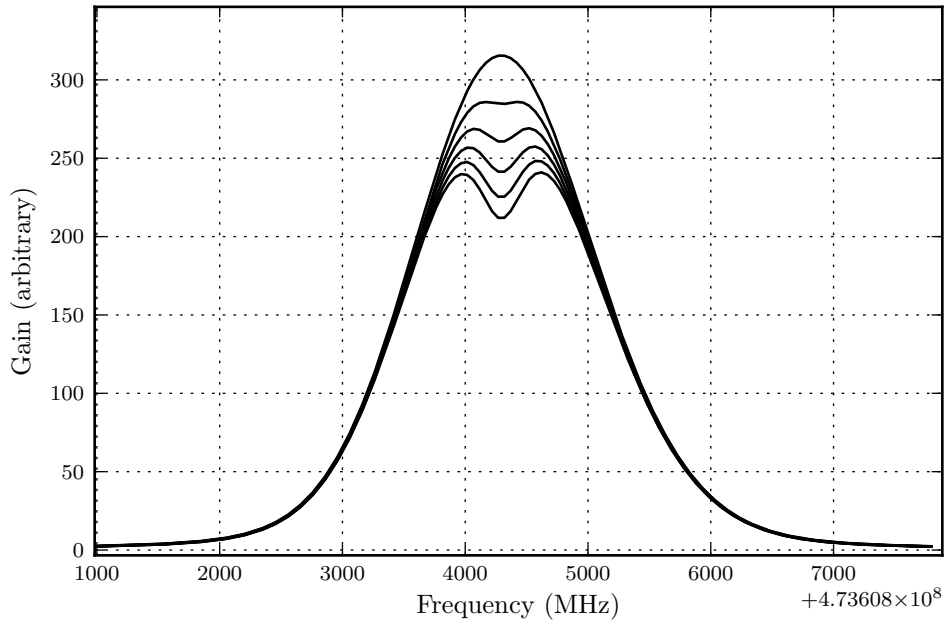


Figure 3.3. A family of gain curves demonstrating greatly exaggerated hole-burning on the natural neon (0.2 mbar) gain curve. $I/I_{sat} = 0 \ldots 0.5$, 5.2 mbar total.

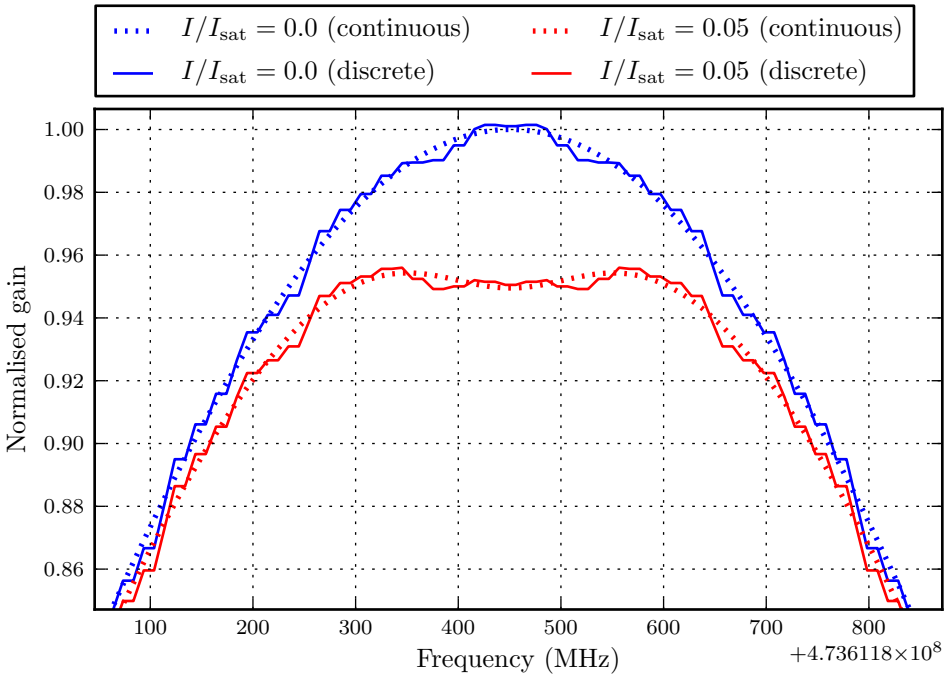


Figure 3.4. Enlarged portion of two gain curves showing discreteisation artefacts.

3.4.1. Split mode

It is possible for clockwise and counter-clockwise beams to operate on different longitudinal mode numbers. We call this *split-mode* operation. We describe split-mode configurations by the *splitting order*, which is the integer multiple of the number of mode spacings separating the two operating beams, disregarding the much smaller Sagnac splitting. In this work we assign the sign of the splitting order relative to the clockwise beam, i.e. a splitting order of -2 indicates the counter-clockwise beam has frequency 2×FSR lower than the clockwise beam.

Operating in split-mode provides a useful method for measuring the FSR of the cavity (see Section 4.5).

3.4.2. The multi-mode threshold

We always operate our ring lasers in the nW power range, and have traditionally filled with a base pressure of 0.2 mbar neon followed by helium to a total pressure between 3 mbar and 8 mbar (giving ratios from 15:1 to 40:1), with the higher pressures for the larger lasers. With low-loss super mirrors this allows a reasonably wide range of operating powers between the threshold of lasing and the *multi-mode threshold*. Under typical operating conditions, multi-mode operation occurs at greater than twice the laser power observed just above the threshold of lasing. Consider a simple description of the behaviour of a laser (frequency f_l, intensity I) operating at the centre of a symmetrical gain curve as the power is increased. To a first approximation we would expect multi-mode operation to begin as soon as the hole burned by the existing mode causes a localised depletion sufficient to increase the gain above threshold for the neighbouring modes on either side. In this case the definition of the multi-mode threshold is

$$g(f_l \pm \mathrm{FSR}, I) \geq g(f_l, I) \tag{3.18}$$

if $g(f_l, I)$ is the composite gain curve. At this point a new mode can start up and begin to challenge the primary mode for gain.

In practice the lasing frequency f_l could be up to $\frac{1}{2}$FSR away from the peak of the un-saturated gain curve f_0. Because of this the multi-mode threshold defined this way has some inherent uncertainty which reduces with FSR.

From calculations of gain curves for commonly used gas mixtures (in particular, non-equal isotope mixtures) and laser parameters it appears that there are two distinct ways through which the multi-mode threshold can arise if following the simple description of Equation 3.18. We describe these as the *low power* and *conventional* multi-mode thresholds.

Low power multi-mode threshold

Consider the gain curve for a laser running with a non-equal isotope mixture (asymmetric gain curve) as the value of I/I_{sat} is increased from zero. Assume that the initial optical frequency corresponds to the maximum of the unsaturated gain curve. As the saturation increases we notice the peak of the gain curve shifts towards higher frequencies (illustrated in Figure 3.5[1]). The derivative of the gain curve at the (initial) operating frequency becomes positive.

Even in the case of relatively large FSR (some tens of MHz), this saturation induced peak shift will cause Equation 3.18 to be satisfied well before any change in concavity (i.e. a hole) occurs. This would suggest that the mode configuration could be affected by very small changes in power. We do not observe this operationally.

In practice multi-mode operation (as opposed to a mode-hop) can not occur unless the primary mode can remain above threshold while the new mode is running[2]. This is only possible if the difference in gain for the primary and new modes corresponds to a doubling of the laser output power. By equating the gain per frequency interval of the saturation-broadened line-width (see Siegman 7.6 [58]) we see that this can happen at surprisingly low values of I/I_{sat} . This is because the fractional change in gain Δg required for a doubling of beam power

$$\Delta g = \frac{\sqrt{1 + 2\,I/I_{\text{sat}}}}{\sqrt{1 + I/I_{\text{sat}}}} - 1 \tag{3.19}$$

can be quite small, for example at $I = 0.1\,I_{\text{sat}}$, the change in gain required for a doubling of the beam power is just 5%. Note that this assumes the two beams are close together in frequency.

[1] In this example, gas mixture was 4.2 mbar total pressure with 0.2 mbar natural neon, temperature 400 K, hole burned at 40.2 MHz.

[2] Note that this neglects the possible case of a change to a configuration with two (or more) modes running, neither of was the primary. This might occur if the gain curve is exactly symmetrical.

The low power multi-mode threshold is then the minimum intensity I such that

$$\frac{\sqrt{1+2\,I/I_{\text{sat}}}}{\sqrt{1+I/I_{\text{sat}}}} \geq \frac{g(f_l + s\,\text{FSR}, I)}{g(f_l, I)} \tag{3.20}$$

where s is an integer.

In practice this condition can not be satisfied given typical laser parameters. Figure 3.6 however shows a theoretical example where a low power multi mode threshold exists, however this was calculated for an unrealistic gas mixture of 0.1 mbar total pressure with 0.02 mbar natural neon at 400 K. FSR was 75 MHz.

Conventional multi-mode threshold

An alternative, and more usual way in which multi-mode operation begins is through hole-burning which causes a localised change in concavity of the gain curve $g(f)$ to occur at the lasing frequency.

In the limit of small FSR, multi-mode operation occurs as soon as the gain curve becomes flat at the lasing frequency f_l, i.e. I such that

$$g'(f_l) = 0, \tag{3.21}$$

where g' is the derivative of the gain curve with respect to frequency. In practice however this must apply over a frequency interval of $1\times$FSR.

The multi-mode threshold can be identified independent of the lasing frequency given a method to calculate gain curves. Above the multi-mode threshold, the first derivative of the gain curve $f'(f)$ has three roots and below the multi-mode threshold it has exactly one. The exact frequency for which the gain curve becomes flat can be identified by finding f such that

$$g'(f) = 0 \text{ and } g''(f) > 0, \tag{3.22}$$

for which there exists only one solution.

Typically we are more interested in simply counting the roots than finding their exact values, so discrete evaluation of gain curves is sufficient. The *LaserGas* class of the *gaincurve* module (Appendix A.1.1, Section 3.3.5) implements a simple uphill-only optimisation routine to find the minimum I/I_{sat} for which the corresponding gain curve satisfies the conditions (Equation 3.22) for the multi-mode threshold.

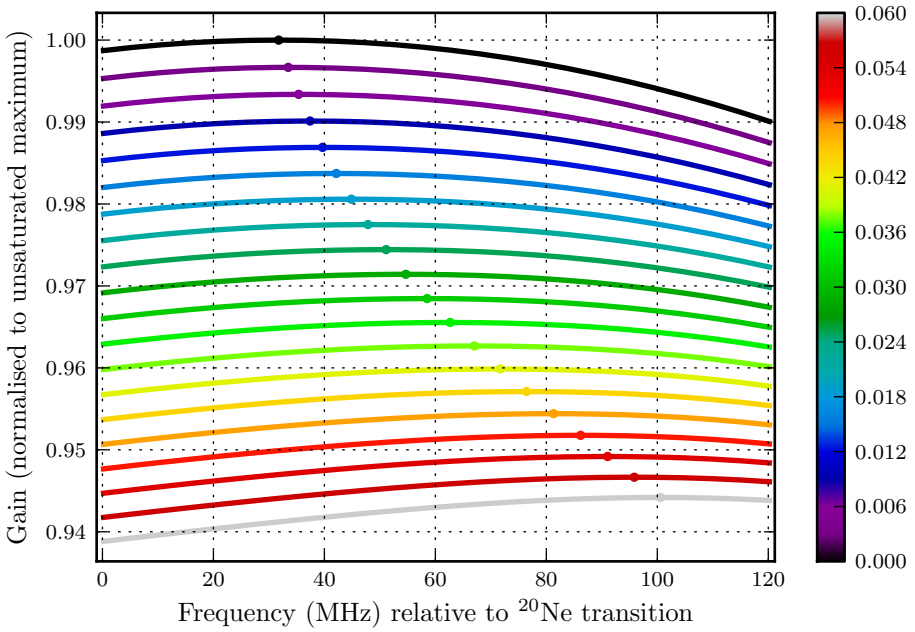


Figure 3.5. Peak-shifting effect in a family of gain curves (colour mapped to I/I_{sat} value). Points have been plotted at the maximums.

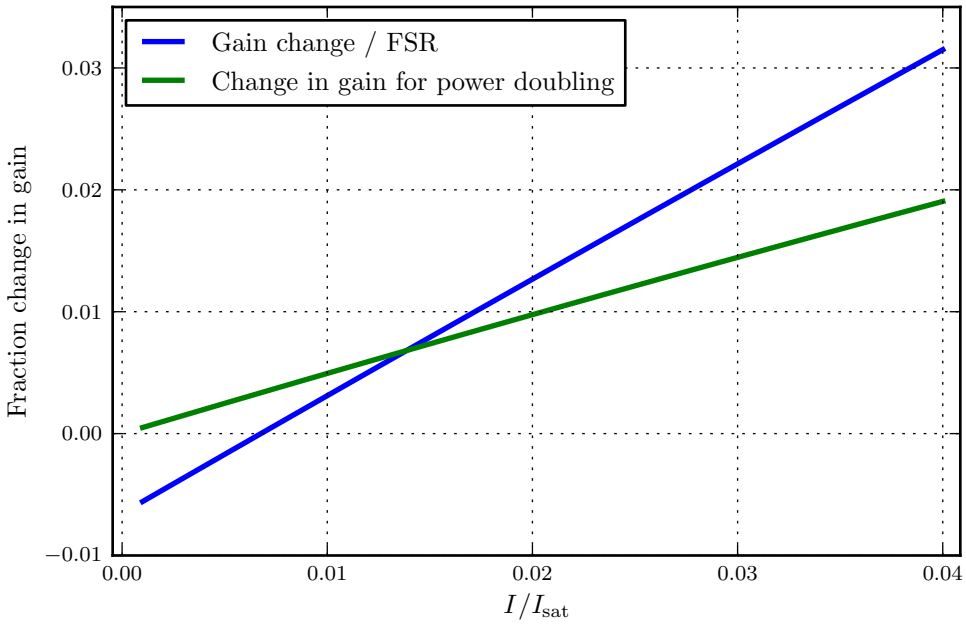


Figure 3.6. Illustration of the low power multi-mode threshold at 0.013 I/I_{sat} .

The optimisation works by iteratively advancing the I/I_{sat} until both crossings are found then backtracking before advancing again with a smaller step size. The FSR is modelled by changing the evaluation resolution. The operating frequency is found initially by finding the maximum of the un-saturated gain curve.

Figure 3.7 shows the pressure and temperature dependence of the value of I/I_{sat} at the multi-mode threshold calculated for natural neon gas mixtures (0.2 mbar neon partial pressure in each case). Based on these results a convenient phenomenological model (Equation 3.23) has been developed. This 4-parameter model was derived by fitting polynomials to the coefficients of exponential fits of the calculated multi-mode thresholds (m) with respect to total pressure p and temperature T,

$$m(p,T) = \exp(d_1 + T\,d_0 + p\,c_1 + p^2\,c_0) \tag{3.23}$$

with the coefficients shown in Table 3.2 for natural and equal ratio neon isotope mixtures at a variety of FSR values.

The variation in these parameters with FSR can in turn be modelled to an accuracy of 2% by simple cubic polynomials, and the respective coefficients are shown in Table 3.3. Overall this model allows the I/I_{sat} value at multi-mode threshold to be predicted quickly given the temperature, pressure and FSR to an accuracy of 5%.

Note that the multi-mode threshold identified in this way does not necessarily represent the true *single mode limit* that would be observed in a real laser. The algorithm does not model the *mode competition* process, in particular the tendency for the operating frequency to shift to neighbouring modes with higher gain and the time constants involved in the establishment of these new modes. In addition, the true single mode limit is expected to vary depending on the exact cavity length (i.e. the offset of the 'comb' of possible modes) at the time of observation[1]. Overall, these effects would all be expected to cause the single mode limit to occur at higher power than calculated here. The multi-mode threshold presented here thus represents a mathematically tractable lower limit to the maximum possible intensity for single mode operation in a real laser.

[1] This is expected to be most significant for the smaller instruments due to their larger mode spacing.

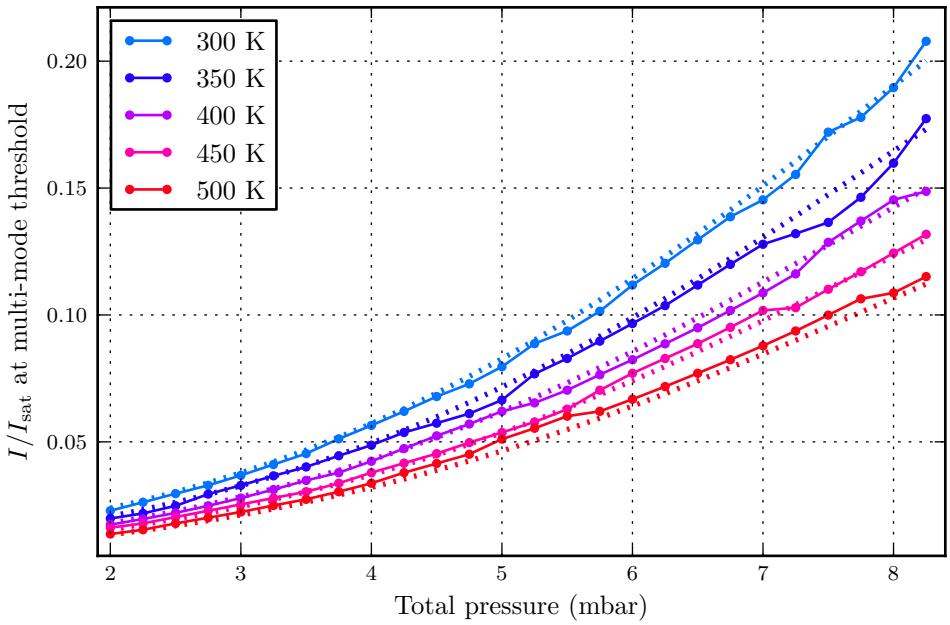


Figure 3.7. Pressure and temperature dependence of the value of I/I_{sat} at the multi-mode threshold. Dashed lines show the phenomenological model. FSR was 50 MHz in this example.

FSR (MHz)	c_0	c_1	d_0	d_1
75	-7.792×10^{-3}	0.3526	-1.869×10^{-3}	-3.431
50	-2.283×10^{-2}	0.5743	-2.893×10^{-3}	-3.925
25	-3.271×10^{-2}	0.7128	-2.126×10^{-3}	-4.756
12.5	-3.409×10^{-2}	0.7305	-1.304×10^{-3}	-5.183

Table 3.2. Coefficients of the phenomenological Equation 3.23 for the multi-mode threshold calculated for various laser parameters.

Coefficent	FSR^0	FSR^1	FSR^2
c_0	-3.489×10^{-2}	-2.128×10^{-5}	5.118×10^{-6}
c_1	7.363×10^{-1}	7.902×10^{-4}	-7.896×10^{-5}
d_0	6.659×10^{-5}	-1.228×10^{-4}	1.290×10^{-6}
d_1	-5.744	4.587×10^{-2}	-1.992×10^{-4}

Table 3.3. Polynomial coefficients for modelling the variation in coefficients c_0, c_1, d_0 and d_1 with FSR for a natural neon isotope mixture.

3.4.3. Establishing a mode configuration

Consider the common situation in which a laser is started with the power level set below the multi-mode threshold. The cavity is initially flooded with a wide spontaneous emission spectrum. A small fraction of this is sufficiently well aligned to recirculate and has optical frequency corresponding to a point on the gain curve above the lasing threshold. This eventually defines the optical frequency of the operating beam.

In practice we observe that immediately after the gain has increased above the lasing threshold, an unstable mode configuration is present on the combined beam. This persists for some considerable time (several minutes with UG-3) and eventually *mode competition* plays out and a stable single mode configuration emerges. This is easily observed on a scanning Fabry-Pérot spectrometer. Conversely, observing the beats between modes present on an un-combined beam (most easily done with a RF spectrum analyser and PMT) we briefly (typically for less than 1 s) observe a complex mode pattern which then quickly disappears.

The splitting order which emerges following a particular restart is random. Should an undesirable configuration (usually anything other than un-split) occur, we must repeat the restart procedure.

In the simplest view of the mode competition process, we consider only the coupling between modes due to hole-burning. This aspect of the process can be simulated quite straightforwardly by assuming some random initial configuration, then applying an iterative process of evaluating the corresponding composite gain curve then multiplying the distribution by this curve until a stable configuration emerges. In

practice however, additional physics is needed to properly describe the mode competition process. A complete view of the process is significantly more involved and difficult to simulate computationally as it involves a number of additional interacting processes:

1. Beating effects between the different modes.
2. Frequency pulling due to backscatter.
3. The degree of spatial overlap between different transverse modes. Note high order transverse modes probably are eliminated very early in the mode competition process. These modes are usually much more widely spaced in frequency than longitudinal modes and as a result the gain discrimination would be much larger.

As an alternative to simulating these many interacting processes, we instead take a more global view of the process, borrowing the terminology of analogue communications theory. In order to gain some insight into how particular mode configurations are established, let us consider the final operating single mode to be a carrier wave with the temporary modes present during startup representing modulation induced sidebands of this carrier.

Amplitude modulation of the carrier travelling around the cavity at velocity c would be expected to appear as modes separated by multiples of the FSR on the RF spectrum of an un-combined beam. That we do not observe this (except for very briefly at the beginning of the mode competition process) suggests that the sidebands can be considered purely the result of phase modulation of the carrier, i.e. a waveform of the general description

$$y(t) = \cos\left(2\pi t + \sum_{i=1}^{N} a_i \sin\left(i\, 2\pi \frac{t}{t_{\max}} + \phi_i\right)\right) \tag{3.24}$$

where $\phi_{1\ldots N}$ and $a_{1\ldots N}$ are the phase and amplitude of each of N cycles of phase modulation. Time t varies from 0 to $t_{\max}$. Figure 3.8 (upper) shows an example phase modulated carrier waveform. In this example the phase modulation is supplied by five randomly generated cycles (amplitudes from 0 to 5).

Figure 3.8 (lower) shows the corresponding mode power distribution $P(n)$, calculated

from the Fourier transform of the phase modulated carrier,

$$P(n) = \mathfrak{Re}\left(\mathcal{F}(y(t))\right)^2 + \mathfrak{Im}\left(\mathcal{F}(y(t))\right)^2, \tag{3.25}$$

and presented with the zero'th mode number corresponding to the centroid of the mode distribution,

$$n_c = \frac{\sum P(n)\, n}{\sum P(n)}. \tag{3.26}$$

An interesting observation from numerical calculations of n_c is that regardless of the values of the random parameters ϕ and a used in generating the phase modulation, the spectral centroid n_c always takes the same value. This is interesting because it indicates that at the point shortly after startup when the amplitude modulation has disappeared, the final mode configuration has already been 'chosen', even though it appears unstable for some considerable time. This indicates that the mode which is to eventually emerge as dominant may be found by calculating the centroid of the spectrum at any time during the mode-competition process.

Note that at any particular time the gain at the centroid does not necessarily correspond to the maximum gain at that instance (as is the case in this example) because the gain at the centroid can be suppressed by the phase modulation (analogous to 'suppressed carrier' modulation schemes). If we reduce the a values which have been used to form the phase modulation, we can illustrate the effect of the mode-competition process during startup. This is shown in Figure 3.9.

Note that Figure 3.9 demonstrates that it is possible for a mode to clearly dominate the other modes for some time, but not eventually emerge as the 'long term' dominant mode. Figure 3.10 shows an example of the mode configuration observed by scanning Fabry-Pérot spectrometer on UG-2 during startup. Qualitatively, we observe similar behaviour. Note that in the experimental plot, the individual modes are not clearly discernible. This is because the data was recorded on UG-2 which has a mode spacing comparable to the resolution available in the scan. No peak finding was applied to the data presented in Figure 3.10.

3.4.4. Mode-coupling

In the simplest picture of a ring laser operating above the multi-mode threshold, each operating mode is independent and produces its own Sagnac signal. Each Sagnac

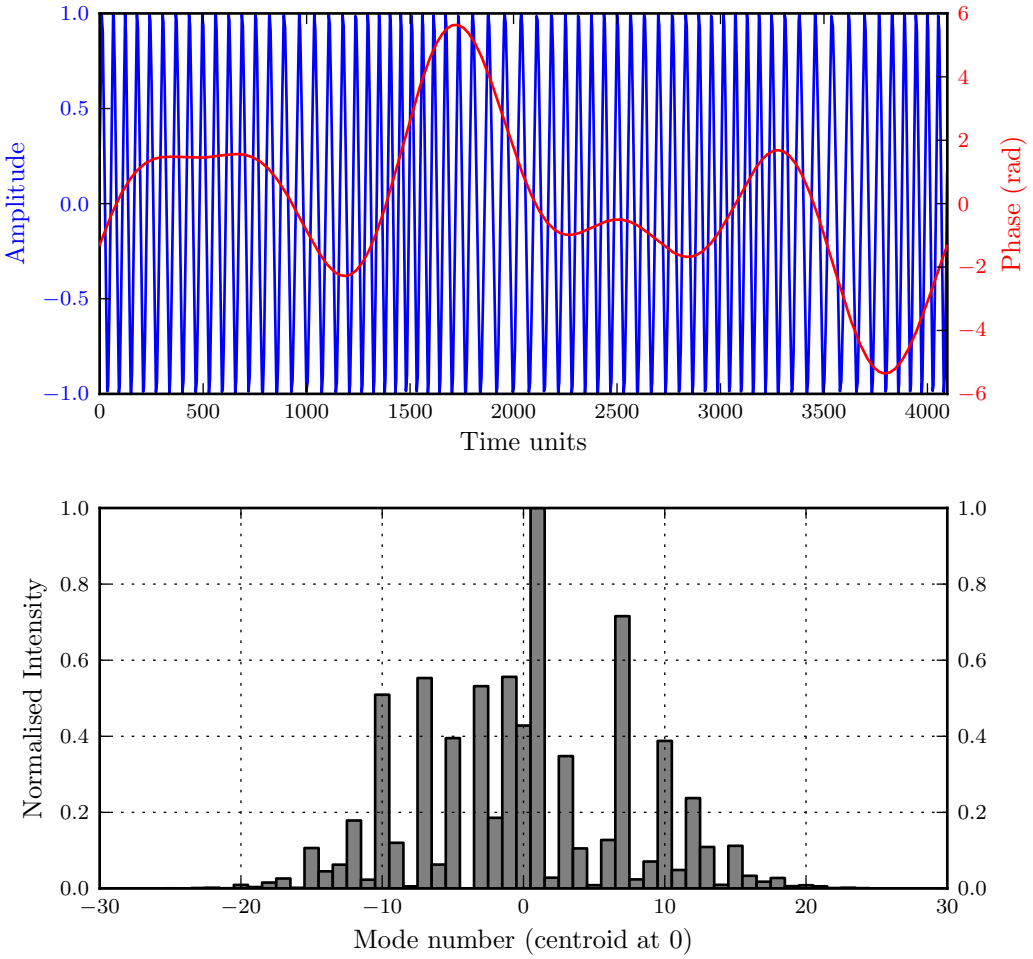


Figure 3.8. An example of a carrier waveform with pure phase modulation (upper subplot), and the corresponding mode power distribution (lower subplot).

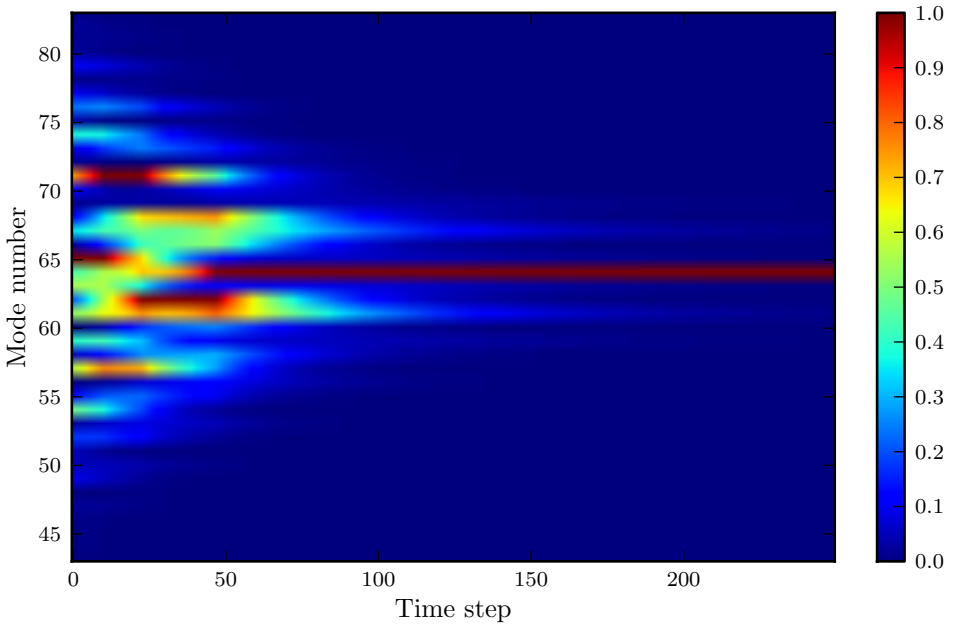


Figure 3.9. Mode configuration during startup modelled by decreasing modulation amplitude (1%/step) of a phase-modulated carrier.

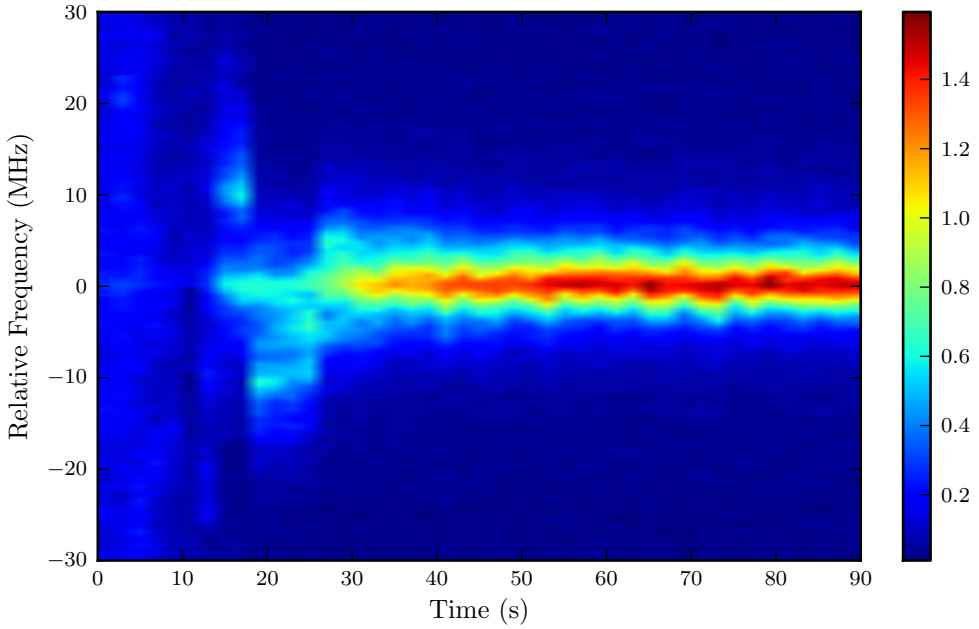


Figure 3.10. Spectrogram style plot of the UG-2 mode configuration during startup.

signal will be a slightly different frequency due to the different wavelengths of the operating modes. An observer monitoring a combined beam would see a complicated but periodic waveform made from the interference of all these signals. In practice the observed waveform is highly unstable because the mode configuration from which it originates is unstable.

In practice we sometimes observe a stable Sagnac beat signal when operating with multiple modes present (identified by Fabry-Pérot spectrometer or RF beat signal). This implies that there must be some fixed phase relationship between the modes. We refer to this as *mode-coupling*.

Mode-coupling occurs spontaneously at specific power levels. When it occurs a Sagnac beat signal with similar frequency and contrast ratio to that achieved in single mode operation can be obtained. The contrast ratio is observed to decrease as beam power increases.

This is understood to be explained by the 'self-mode-locking' phenomena. Self-mode-locking has been observed in gas lasers since the 1960s [80–83]. The process has been described by Lamb (Section 14) [84] as a perturbation in the laser oscillation which occurs via the injection of an 'external' tone that arises through the third-order polarisation of the gain medium.

The conditions under which mode coupling occurs in ring lasers is currently a topic of research being addressed by Holdaway.

3.5. Splitting statistics

In the description of the start-up process detailed in Section 3.4.3 we noted that the 'seed' which defines the mode that (following a lengthy competition process) will oscillate is defined very quickly in the start-up process.

Note that for a given start-up of UG-3, the final modes are rarely split by more than 3×FSR. Over this range of frequencies a typical composite gain curve is essentially flat. We would expect from this that stable operation should be possible over several hundred MHz. However we do not observe this in practice. While the single cavity mode that emerges following any particular restart is random, it is chosen from a particular statistical distribution. Because we observe this distribution to be much

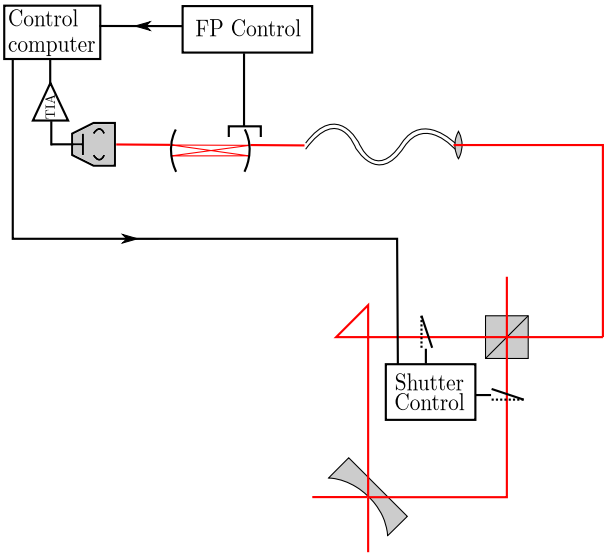


Figure 3.11. Setup (simplified) for alternately measuring UG-3 mode configuration of each beam.

narrower than the regular composite gain curve, we can use the statistical distribution of operating modes to define a line-shape for the 'seed' transition.

It is possible to record the mode configurations by Fabry-Pérot interferometer for each beam in quick succession, shortly after start-up. This can be used to establish the splitting order. The statistical distribution of splitting orders from many restarts can then be established. Note that using the splitting orders assumes that there is no interaction in the process which 'chooses' the operating mode in each direction. This is quite reasonable, given the small (not directly observable) modulation on the un-combined beams (indicative of any backscatter) for UG-3.

3.5.1. Experiment

The basic experimental setup used here is illustrated in Figure 3.11. A combined beam is focused and carefully aligned using a precision fibre launcher (Newport F-91-C1) into a single mode optical fibre. The fibre emerges on an optical bench where the light is focused by a beam expander into a scanning Fabry-Pérot interferometer (Coherent model 216, spherical configuration).

Either of the individual beams can be selected quickly by opening and closing shutters

installed prior to the beam combiner. The shutters were designed for this experiment and make use of micro servo-motors which move light wooden levers (∼100 mm in length) into the path of the beam. Both servo motors are controlled by a Atmal ATMega168 micro controller, using a C program written by the author to produce the pulse-width-modulated signals for the servo motors in response to commands received via a RS-232 connection. Commands specifying the desired lever angle and motor identifier and can be sent from an associated LabVIEW program. This approach was found to be highly reliable and able to operate without causing any detectable vibration[1] of the sensitive beam combiner and fibre launching optics.

In these experiments we are only concerned with recording the locations of the peaks in the Fabry-Pérot scan. These experiments were done with UG-3 and the raw spectra were filtered using the technique described in Section 6.6.5. The individual scans were made over 300 MHz optical frequency range at a rate of 50 Hz with 10 scans acquired and averaged for each measurement. The raw acquisition rate was 500 kHz and scans were triggered by an analogue threshold detector from the Fabry-Pérot controller reference signal in order to eliminate fly-back artefacts.

Overall the noise level was low enough to allow the peaks to be identified during acquisition by a simple LabVIEW peak finding routine[2]. This routine works by performing a least-squares quadratic fit with a specified width to blocks of data. A series of tests are then performed on the coefficients to see whether they meet the criteria for a peak, these are:

1. Whether each parabola is at a local maxima.
2. Whether the parabola is concave up (found from the sign of the quadratic coefficient).
3. Whether the interpolated data value at the peak is above a designated noise threshold.

This technique is useful because it effectively interpolates between the data points while finding peaks. The relative peak frequencies and heights are then returned as floating point numbers from which the centroid can be calculated (see Equation 3.26).

[1] This was not the case with earlier attempts where micro solenoids were used.

[2] See LabVIEW documentation http://zone.ni.com/devzone/cda/tut/p/id/3770 for more information on this routine

One problem with the optical configuration shown in Figure 3.11 is that there is a large imbalance in the beam powers. While the beam splitters have a nominal 50% splitting ratio, this only applies to circularly polarised light. In practice the intensity of the reflected beams always exceeds the transmitted beams (typically with ratio of 8:2, varying slightly with different optics) so the counter clockwise beam is brighter and thus has a lower noise level[1]. For this reason we always measure the centroid of the clockwise beam both before and after measuring the counter-clockwise beam, averaging both measurements.

An advantage of measuring the centroids of the counter-clockwise beam twice is that we can estimate the rate of thermal drift in the Fabry-Pérot. The laboratory is reasonably well thermally insulated and the shutters only require about 200 ms to operate. For these reasons the rate of thermal drift is usually negligible unless the laboratory has been recently disturbed. Nevertheless, comparing the two measured centroids provides a useful diagnostic and can be used to eliminate erroneous measurements.

Restarting of the laser was achieved by interrupting the RF power to the discharge under software control. Following interruption we introduce a variable time delay, followed by a sharp rise to maximum power which must be sustained for around 70 ms. We then slew the power back to the operating level linearly over 10 ms. This restart procedure is necessary because of the limited bandwidth of the RF transmitter output control. Faster slewing of the control voltage tends to cause the discharge to stop. This restart procedure is illustrated in Figure 4.18. It is possible that optimisation of the electrodes and impedance matching circuit may allow the time delays to be reduced.

3.5.2. Analysis

Figure 3.12 shows a histogram of the identified splitting orders observed after 13 740 restarts which were made over several days. In this example the orientation of the splitting (i.e. the sign of the splitting order) is of no consequence, we are really only interested in the difference in mode numbers between the beams.

[1] This problem could be eliminated by using a $\lambda/2$ plate to rotate the polarisation of the clockwise beam before the beam combiner, but this was not available at the time of the experiment.

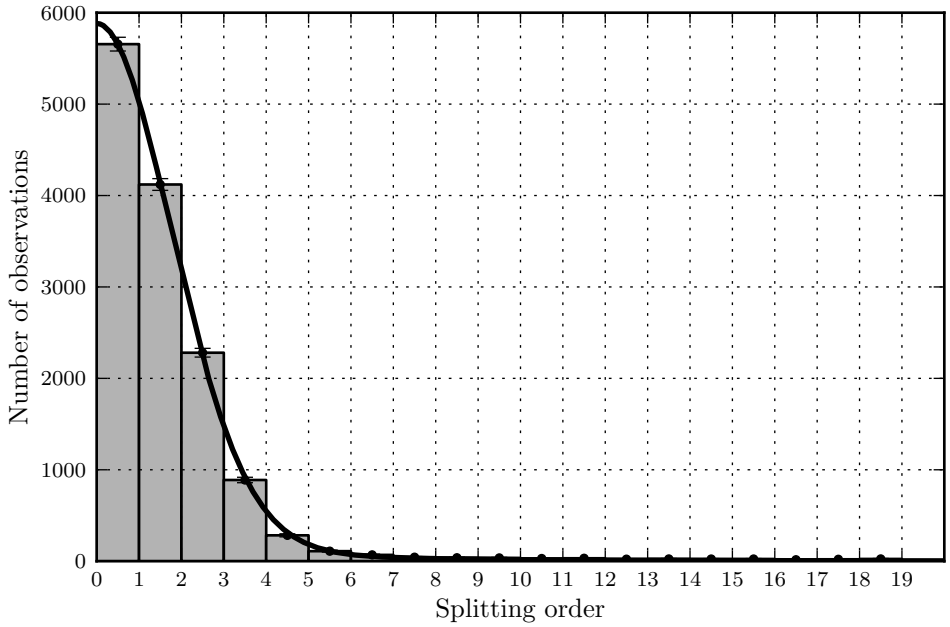


Figure 3.12. Histogram of the identified splitting orders observed after many restarts with a Lorentzian fit. Error bars are from the 1-sample t-test.

Because the Fabry-Pérot is subject to thermal drift, it is possible for the pair of modes to move along the scan until one mode 'wraps around' and appears on the opposite extreme end of the scan and as a result the splitting appears to be slightly less than 300 MHz (the Fabry-Pérot scan range). This situation is automatically disregarded by the software[1] and as a result there is a slight statistical bias towards low order modes which are less likely to be subject to this problem. This has been corrected for in the data set presented in Figure 3.12 by adjusting the counts of the number of observations N for each splitting order S,

$$N'(S) = N(S)\left(1 + \frac{S\,\mathrm{FSR}}{300}\right). \tag{3.27}$$

From Figure 3.12 we immediately see that the distribution is much narrower than the gain curve which under typical gas parameters is essentially flat over several hundred MHz.

Consider the 'seed' photons which eventually define the operating modes in each direction. It is reasonable to assume these are drawn from a statistical distribution which originates as a spectral line, broadened as usual by homogeneous and inhomogeneous processes. By applying a least-squares fit of the Voigt function (Equation 3.17) to our observed distribution of splitting orders we can determine the Lorentzian and Gaussian widths associated with the distribution. This is shown in Figure 3.12, here the fit is very good with with the coefficient of determination $R^2 = 0.9998$.

The observed distribution of splitting orders comes from the difference between two random variables (the separate distributions for each beam). If these are statistically independent then the splitting distribution is the convolution of the distribution of a single beam with itself. By observation of the behaviour of the component curves under convolution, we can recover their original widths. The Lorentzian width is halved while the Gaussian width is reduced by the factor $\frac{1}{\sqrt{2}}$.

Note that in this analysis we are considering the Voigt function as a probability density distribution. It is important to note that this function has infinite variance which implies a finite probability of observing arbitrarily large splitting orders. This

[1] Complexities arise in discerning this situation, particularly if a mode is partially 'wrapped around'. Additionally, it is difficult to calibrate the scan length to exactly correspond to the 300 MHz mode spacing.

however is not possible due to the finite width of the gain curve. This analysis is only strictly valid for splitting orders where there is minimal gain discrimination between the modes, however this is easily valid for splitting orders up to about 100. Small but finite probability at large splitting orders (at least 10 observations at $S = 20$) have been observed in this experiment.

From the fit to our data we find the Lorentzian width to be 0.653 MHz. Correspondingly, from the transition probability (from the NIST Atomic Spectra Database [57]) we would expect the lower limit for the natural line-width to be 0.54 MHz. A slightly higher value is expected because the true natural line-width also depends on the rate of any non-radiative decay processes.

This suggests that any collisional broadening effects are minor. This in turn suggests that the final operating mode is determined very early in the start-up process, i.e. before collisions have occurred. This provides some validation to our general description of the laser start-up process discussed in Section 3.4.3. It also demonstrates an interesting technique for finding the natural line-width.

From the fit to our data we find the Gaussian line-width to be 11.1 MHz. Additional analysis is needed to determine how this width is established.

Consider the laser very shortly after startup. At this point the modes are 'growing' freely and independently and saturation has not yet had time to take effect. The distribution from which the final operating mode is selected should be determined by the corresponding distribution of the relative rates of growth of the modes. Determining the observed Gaussian width may be possible by modelling this aspect of the start-up process by computer simulation. This is left as further work.

3.6. Gas temperature

One variable which has been mostly overlooked in earlier work on ring laser gyros and the helium-neon system in general is the gas temperature. However, as we have seen in Section 3.3, changes in gas temperature can significantly affect the shape of the gain curve and thus the saturation behaviour. The latter occurs because the saturation intensity is inversely proportional to the effective recovery time of the transition which in turn depends on temperature.

In earlier work by the author [85], gas temperature was investigated by measuring the change in pressure in a small laser cavity when the discharge was switched on. Unfortunately, further investigations cast some doubt on this method of temperature measurement because the small pressure changes are difficult to observe in the presence of RF interference which inevitably accompanies a RF excited plasma. RF interference is also a problem for any method of finding the temperature of the gain tube using a thermocouple or other electronic temperature sensor. A possible solution to this type of measurement might be building an alcohol[1] thermometer into the side of the gain tube.

3.6.1. Variation with RF power

Hurst[2] has suggested a technique for estimating the gas temperature from RF power measurements. The basic idea is to measure the wall temperature T_{wall} at thermal equilibrium for a sample gain tube with resistive heating elements placed inside. The heating elements are powered by a DC power supply which allows the power dissipated to be easily found from the current and resistance. This experiment was carried out by Hurst with a 6 mm internal diameter gain tube (from UG-2). The relationship between power absorbed per mm of gain tube $P_{\text{W/mm}}$ was found to be well modelled by,

$$P_{\text{W/mm}} = T_{\text{wall}}^2\, 7.45 \times 10^{-7} + T_{\text{wall}}\, 2.25845 \times 10^{-4}. \tag{3.28}$$

Now given the relationship between absorbed power and wall temperature, a simple model could be used to find the radial heat flow from the core of the plasma to the tube wall. This was done by dividing the total radius R into radial segments of length s. First assuming the total power P is uniformly distributed through these segments

$$P_{\text{dis}}(r) = P\frac{(r+s)^2 - (r-s)^2}{R^2}. \tag{3.29}$$

The total temperature difference could then be found by integrating the power flowing per unit area $A(r)$ with thermal conductivity $C(T)$

$$\Delta T = \int_0^R \frac{P_{\text{dis}}(r+s) - P_{\text{dis}}(r)}{A(r)C(T)}\mathrm{d}r \tag{3.30}$$

[1] It would be necessary for the working fluid of the thermometer be non conductive.
[2] Via private communication.

which when evaluated yields

$$\Delta T = \frac{P}{L(3.41 \times 10^{-6}\, T_{\text{wall}} + 0.001006)} \tag{3.31}$$

where L is the length along the plasma in mm.

Note that the thermal conductivity is a function of temperature and gas composition, so Equation 3.30 should be evaluated starting from the known temperature at the tube wall. The thermal conductivity of helium in $\mathrm{W\,K^{-1}\,m^{-1}}$ can be linearised and found independent of pressure from

$$C(T) = 3.04 \times 10^{-4}\, T + 0.0613 \tag{3.32}$$

to within 5% [86].

Lastly, it is necessary to account for the variations in power dissipation along the length of the tube and the 'smoothing' afforded to this distribution by the thermal conductivity of the glass. In our ring lasers we typically use a system of three equally spaced electrodes with the RF excitation applied to the middle electrode and the outer electrodes grounded. This leads to two lobes of plasma being established and a rather unusual distribution along the length of the tube as illustrated in Figure 3.13.

Hurst has approached this final step by generating a piecewise function to represent a distribution of brightness roughly matching the observed 'double lobe' distribution (Figure 3.14). This function is used to distribute the total power in sections along the length of the tube. The thermal conductivity of the glass is modelled by applying a exponential filter with $1/e$ distance of 20 mm.

Summary of this approach

The overall dependence of plasma temperature on absorbed power predicted by this analysis is shown in Figure 3.13 over the typical range of operating powers. From this it is clear that plasma temperatures can vary over a wide range and this should be taken into account in any work involving the gain process.

Many assumptions have been made in this analysis, and it is therefore difficult to assign any reasonable uncertainty to the results, however they do represent a useful estimate for further investigations into gain processes.

This approach assumes that the visible brightness of the plasma is proportional to the

absorbed power. A linear relationship here was shown by the author in earlier work [85], in which RF excitation was shown to produce light over the visible spectrum with constant slope efficiency of 0.38%.

Another assumption is that the RF power (measured using an analogue RF power meter) accurately represents the power absorbed by the plasma, i.e. that the impedance matching circuit is 100% efficient. It is known from operational experience that components in the matching circuit can become quite hot, this heating is usually minimised when the circuit is well matched, however in practice it is not possible to perfectly match the impedances and some power is always radiated.

This approach also ignores any potential convection (or otherwise induced) currents in the plasma, which would result in a more efficient transfer of heat through the plasma. The phenomena of Langmuir flow (discussed further in Section 3.8.2) is well known in gas lasers, and would be expected to increase the rate of heat flow through the plasma.

3.6.2. Variation with hydrogen accumulation

We observe that the RF power required to maintain constant laser power increases over time. Our understanding, backed by spectroscopic measurements [85], is that this is due to a build up of hydrogen which out-gasses from the stainless steel tube walls and decreases the efficiency of the pumping process by providing additional de-excitation pathways. This results in an increased excitation power, and thus the power absorbed by the plasma which then causes an increase in temperature. The increase in temperature reduces the saturation intensity (see Section 3.7.2) which in turn reduces the multi-mode threshold. This leads to a long term behaviour which can be summarised:

1. The accumulation of hydrogen reduces the efficiency of the neon excitation process.
2. The servo control system increases RF power to compensate.
3. At higher RF power, the gas temperature increases.
4. The absolute gain available decreases because the total available neon atoms are spread over a larger frequency range, leaving decreased population at the

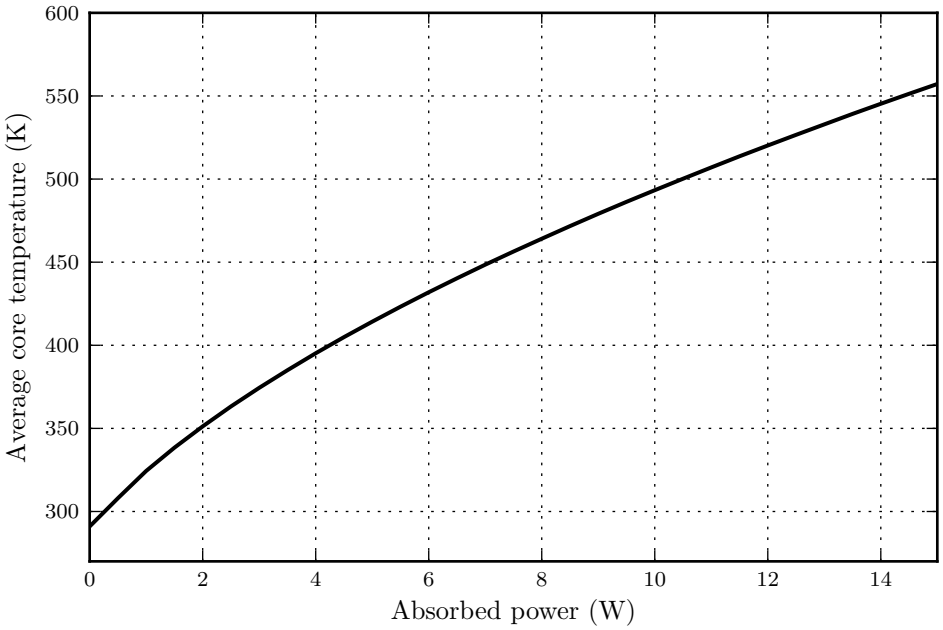


Figure 3.13. Average core temperature for a range of common RF power levels.

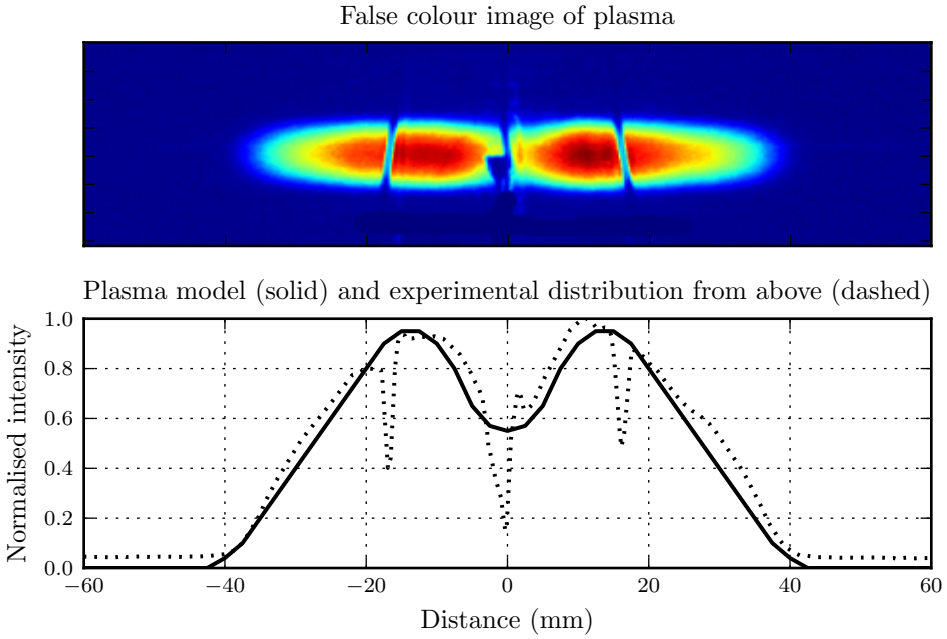


Figure 3.14. Typical plasma distribution from a photograph and the distribution model. Note the vertical lines in the image are the electrodes.

lasing frequency. This further compounds the increase in RF power.

5. The temperature gradient from the core of the plasma to the wall is now steeper. If there are any asymmetries in the discharge this would be expected to increase the rate of any associated gas flows. These gas flows can cause null shifts which result in a long term Sagnac frequency drift.

6. As the temperature increases, the multi-mode threshold decreases (a decrease of 1%/K is typical). Eventually the desired operating power is now above the multi-mode threshold, and the laser must be refilled with fresh gas.

3.7. Saturation intensity measurement

Because of the inconsistencies in measurements of the saturation intensity available in existing literature, an experiment was undertaken to measure the saturation intensity and determine the relationship to gas pressure, temperature and mixture. The basic procedure was to find the intensity (I) corresponding to multi-mode threshold experimentally, then establish the corresponding value of I/I_{sat} from gain curve calculations and solve for the saturation intensity.

3.7.1. Experiment

Establishing the multi-mode threshold experimentally was done by slowly[1] increasing the laser power from the single mode level while taking 'snapshots' of the mode configuration (typically every 1 s) observed on a scanning Fabry-Pérot interferometer while also recording the beam power. The multi-mode threshold was identified as the first instance at which a second mode started up and began to challenge the original mode, reducing its power.

Measurements were obtained for different gas pressures (between 1.5 mbar and 5.7 mbar) and mixtures (5%, 7%, 8.5% natural neon) on both UG-2 and UG-3. The majority of the measurements were made with UG-2. Making a new gas fill and performing the measurements typically took around 1 day.

[1] It is necessary to increase the power as slowly as is practical so as to not confuse a mode jump with true multi-mode operation. In this experiment the power was increased by adjusting integration time of the servo controller to the maximum setting (10 s). Typically the power would increase from typical operating level to the multi-mode level over around 10 minutes.

With UG-2, identifying the modes is difficult because the mode spacing is only 2.47 MHz which is small relative to the Fabry-Pérot peak width. With UG-2 it was only possible to achieve a finesse[1] of around 300 with careful alignment and the use of angled spherical lenses to correct for the astigmatism of the UG-2 beam and minimise the spot size. The finesse was found to be very sensitive to misalignment. In order to obtain the multi-mode thresholds for UG-2 it was necessary to regularly record spectra and laser power, then process these many spectra (of the order 50,000) off-line, looking for multiple peaks. With UG-3 we did not have the same difficulty as with UG-2, with the larger mode spacing the multi-mode threshold could simply be identified in real time by the experimenter.

In the experiments with UG-2, the Fabry-Pérot scans were recorded by a photomultiplier tube followed by a trans-impedance amplifier and low pass filter with ∼1500 Hz pass band. The filter was necessary to remove the noise discussed in Section 6.6.5[2]. Beam powers were taken from a calibrated photomultiplier tube monitoring a single beam.

The method used to identify multi-mode operation can be summarised:

1. From a peak recorded when the laser was known to be operating in single mode, establish the Lorentzian width F of the peak.

2. Re-sample and normalise the recorded spectra so that the maximum value falls at $f = 0$.

3. Apply a least-squares fit for each recorded spectra to the model of the power signal $y(f)$, i.e.

$$y(f) = \left(\frac{\left(1 + \left(\frac{f}{F}\right)^2\right)^{-1} + h\left(1 + \left(\frac{f - S\,\mathrm{FSR}}{F}\right)^2\right)^{-1}}{1 + h\left(1 + \left(\frac{f}{F}\right)^2\right)^{-1}} \right)^2 \tag{3.33}$$

   (chosen to match the filtered Fabry-Pérot lineshape) in order to obtain the integer S and the relative power of the second beam h. Note that the denominator of Equation 3.33 ensures the maximum value is always unity.

[1] Taken from the ratio of the FSR to the bandwidth of the Fabry-Pérot cavity.

[2] Note that the digital filtering technique discussed in Section 6.6.5 had not been developed at the time of the experiment.

4. Given all the measurements of splitting order, apply filters to eliminate erroneous measurements. These involved removing all measurements such that the regression coefficient

$$R^2 > 0.8 \tag{3.34}$$

and in which the identified integer S satisfies

$$|\text{round}(S) - S| < 0.25, \tag{3.35}$$

i.e. where the identified number of modes separating the peaks is clearly not an integer multiple of the FSR, and where the power ratio

$$h > 0.1\,. \tag{3.36}$$

5. Manually identify a series of successive (non zero) splitting orders and find the power level corresponding to the beginning of the series.

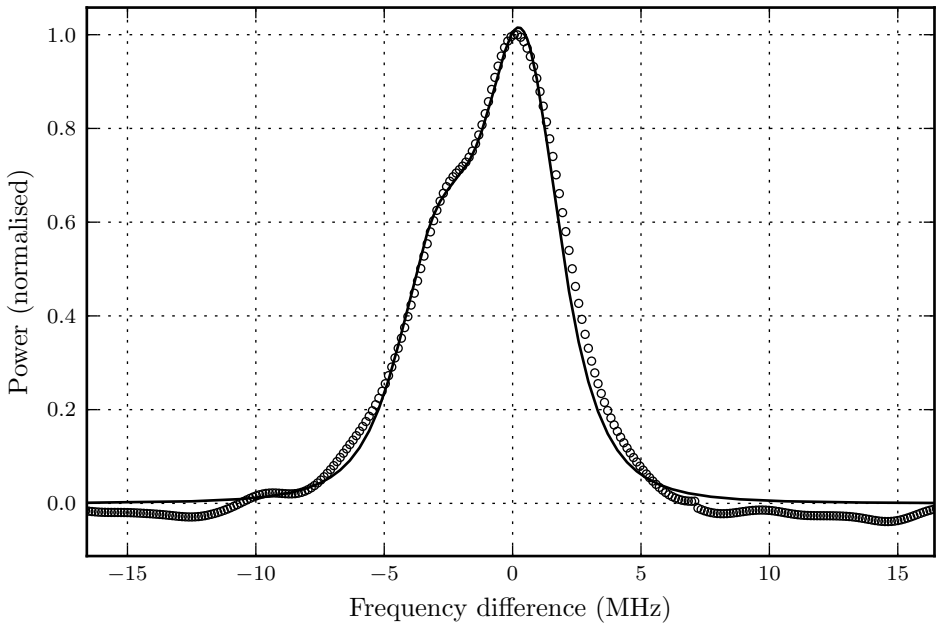


Figure 3.15. An example spectra recorded on UG-2 with fitted curve (Equation 3.33).

The intra-cavity intensity at the gain tube was calculated from the beam power given a measured mirror transmission of 0.147×10^{-6} (for UG-2, south east corner) and calculated spot size at the gain tube (according to Siegman's definition, $2.62 \times 10^{-2}\,\text{cm}^2$

for UG-2). Measurement of the transmission of mirrors has been done routinely using a ∼10 mW helium-neon laser, laser power meter (for the incident beam) and calibrated photomultipler (for the transmitted beam). Transmission of mirrors has been found to vary across the surface of the mirror, but with variation less that 3%.

3.7.2. Analysis

Given the calculated value of I/I_{sat} for each gas fill (see Section 3.4.2) and experimental measurements of the intra-cavity intensity at multi-mode threshold we can calculate the saturation intensity. Ideally we would like to relate this to the gas parameters, i.e. pressure, temperature and gas mixture. Thus we require a parameter to use for the fitting which takes these into account in a realistic way. If we recall Equation 3.13, we see that the effective collision time is a factor determining the saturation intensity. In an ideal gas at constant pressure, the mean speed of the atoms varies proportional to $\sqrt{T}$, while the density is proportional to $1/T$. The collision frequency[1] is therefore proportional to $1/\sqrt{T}$. Similarly, the density is proportional to pressure p suggesting that $p/\sqrt{T}$ would be a suitable parameter to use.

Ideally we would also like to account for variations in neon partial pressure, however in our measurements no statistically significant difference was observed in the saturation intensity as neon partial pressure was varied from 5% to 8.5% while maintaining the same total pressure (2.8 mbar).

Figure 3.16 shows the saturation intensity as a function of $p/\sqrt{T}$ measured on UG-2. A linear least-squares fit to this data allows us to present a phenomenological model of the saturation intensity,

$$I_{\text{sat}} = \frac{124.5\, p}{\sqrt{T}} + 10.17 \tag{3.37}$$

in W cm^{-2}. This model can be used for calculating gain curves and is available in the *physconst* module, *saturation_ intensity()* function (see Section A.2.1).

Based on the measured values, this model can be considered accurate to only 17%. Comparing the saturation intensities to equivalent values[2] using the model (Equation 3.14) derived from the data presented by Smith [68, 73] we note a similar gradi-

[1] Note that counter-intuitively, the collision frequency in a gas at constant pressure reduces as temperature increases.

[2] Here we have generated a point for each of our measured pressures, this effectively represents

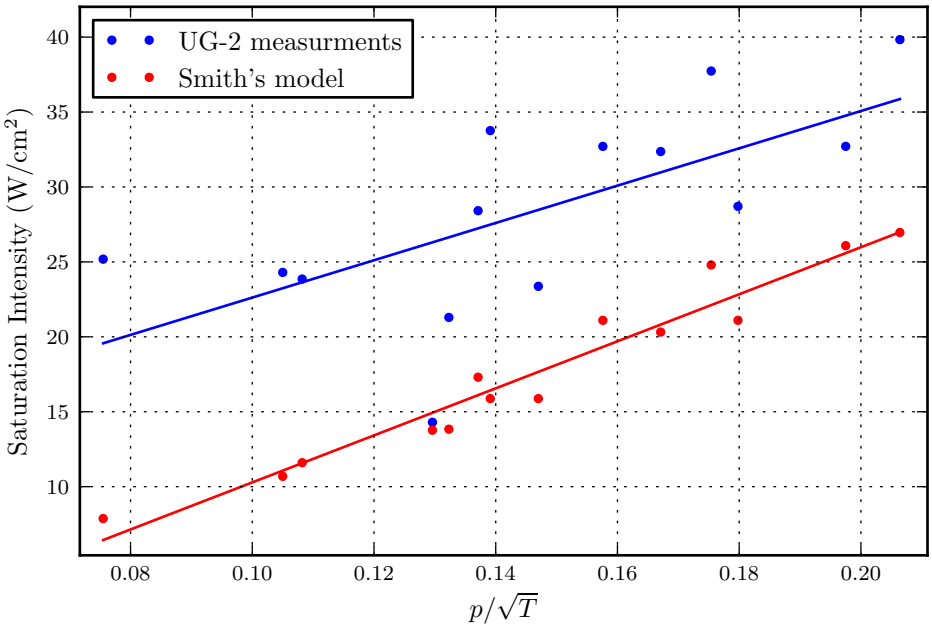


Figure 3.16. Saturation intensity measurements from UG-2, compared with a equivalent measurements calculated from a model (Equation 3.14) derived from the data presented by Smith [68, 73].

ent, but constant offset ($\sim$15 W cm^{-2}) in each case. Recognising that in these earlier measurements no account of temperature variations was made, that there was an ambiguity in the effective area, and that there were experimental differences (He^3, DC excitation) our results are not significantly different. Furthermore, recall from our discussion in Section 3.4.2 that a number of effects exist which would be expected to reduce our theoretical multi-mode threshold values compared to what would be observed in practice. This would have the effect of reducing our experimentally derived saturation intensities which would bring them closer to Smith's results.

One source of error in these measurements is the estimation of plasma temperature. As we have seen in Section 3.4.2, temperature changes of around 100 K can cause changes in the observed multi-mode threshold of up to 30% depending on discharge parameters. At present we can estimate the temperature only indirectly from RF

a working assumption that the relationship between pressure and temperature is the same for both data sets.

power measurements. As we have seen in Section 3.6, there are a number of assumptions in this estimation which mean it is difficult to estimate a reasonable uncertainty, although we note that the initial RF power measurements are at best accurate to around 5%.

An additional source of error stems from the difficulties in estimating the multi-mode threshold from Fabry-Pérot scans. In reality, mode competition makes this a dynamic process; it is possible for a second mode to start up and for its power to briefly exceed the power of the primary mode, but then quickly collapse. In analysing the data it is necessary to choose reasonable criteria for eliminating measurements where the mode configuration can not be determined; however some erroneous measurements will always be missed. This can lead to an error in identifying the true multi-mode threshold from what might be better characterised as a transient phenomena.

3.8. Gain medium refractive effects

3.8.1. Dispersion

In the earlier analysis of laser gain, we have considered only the real part of the gain curve. In a complete treatment the gain curve of course has an imaginary component, which corresponds physically to dispersion in the gain media which inevitably accompanies gain. It is possible to account for the dispersion quite straightforwardly without resorting to the complete treatment by making use of the Kramers-Kronig relations.

Close to the centre of the lasing transition the dispersion varies approximately linearly with optical frequency. This pulls the modes together, reducing the effective Sagnac splitting and thus causing a lower than expected Sagnac frequency.

Usually the frequency pulling will only be a small fraction of the FSR. The easiest way to calculate the size of the frequency pulling is to approximate the composite gain curve by a Lorentzian of width l_w, following Seigman Equation 12.18 [58]. The pulling of the lasing frequency Δf_l is

$$\frac{\Delta f_l}{f_l} \approx \frac{G_0 \,\mathrm{FSR}}{2\pi l_w}. \tag{3.38}$$

Taking UG-2 as an extreme example (single pass gain G_0 of 800×10^{-6}, FSR of

2.47 MHz, assuming gain curve width of 1.800 GHz we obtain $\Delta f_l / f_l$ of 1.7×10^{-7}. This equates to a reduction in the Sagnac frequency of just 0.3 mHz.

At the present stage of ring laser development it has not been necessary to apply a correction for dispersion as part of routine observations because the noise level has always been well above the required correction. Even as mirror technology improves, this is likely to remain the case for some time because the single pass gain will reduce correspondingly.

3.8.2. Gas flows

The phenomena of Langmuir flow is well known. Its effect in ring lasers was first described by Aronowitz in 1967 [16].

By way of a general description of this process, consider the effect of the electric field used to excite the plasma on the electrons and ions inside the gain tube. The ions and electrons are in equal numbers in a uniformly ionised gas and each move in opposite directions along the electric field lines. When an ion collides with the tube wall (a process we generally try to encourage in order to maximise gain) its momentum will be transferred to the wall and not the gas directly.

A negative charge sheath[1] builds up on the wall which repels the electrons, allowing only the fastest electrons to reach the wall. Because of their mass, the positive ions give up much more momentum to the walls than the electrons. This causes a layer near the wall (thickness approximately one ion free path length) in which greater momentum is given to the electrons than the ions. If the plasma is to remain overall electrically neutral, the electron flow must be compensated by an opposite flow of ions parallel to the wall.

Podgorski and Aronowitz [13, 16] have reported gas flow rates approaching 100 mm s^{-1} in a DC discharge. They report a linear relationship of 2.8 mm s^{-1} per 1 mA of unbalanced current. They also note that while gas flow rate is a linear function of current in a DC discharge, laser power is not and therefore a non-linear power dependent shift should be expected.

[1] The existence of this sheath and resulting higher gain can sometimes be exploited, for example Schuöcker [87] has reported on experiments with very narrow diameter gain tubes where the extension of the sheath is comparable to the tube diameter.

We use RF excitation to minimise any preference for flow in any particular direction. However the RF fields are easily influenced by the particularities of the electrical grounding of the surrounding equipment, and the electrodes are only nominally symmetric. This inevitably leads to some asymmetry in the fields inside the plasma which would be expected to result in a net flow of gas.

There are two mechanisms through which gas flow can affect the rotation rate measurement:

1. Because of the refractive index of the gas, Fresnel drag can cause a null-shift in the Sagnac frequency, analogous to the description in Section 2.5.2 for a spinning glass wheel or flowing liquid. However, because of the low refractive index of the gas (see Section 3.9) this is a very small effect and would require flow rates of tens to hundreds of meters per second to cause a noticeable shift.

2. Gas flow can cause an offset in the Doppler broadened velocity distribution. This can move the operating frequency and in turn introduce an offset in the Sagnac frequency. This can be simulated with the *gaincurve* module (Section 3.3.5), which does so by simply adding a Doppler offset $\frac{v}{\lambda}$ to the Gaussian curve when constructing a Voigt gain curve. Overall this is a much more significant effect and can cause an offset in the Sagnac frequency of a few mHz with flows of just a few $\mathrm{mm\,s^{-1}}$.

3.9. Cold-cavity refractive effects

The exact value of the refractive index inside the laser cavity at the frequency of the lasing transition (n) is important if the exact perimeter is to be determined from the mode spacing.

The majority of the gas in the laser is helium. This is convenient because the refractive index of helium is very well known. It can be calculated directly using *ab initio calculations* [88] and for this reason can serve as a theory-based refractive index standard for high-accuracy interferometric length measurements.

The refractive index of helium at 632.8 nm, 101 325 Pa and 293 K is 1.00003242600 $\pm$ 8×10^{-11} [88]. It is possible to find this as a function of pressure given precise measurements of the Virial coefficient, however with the pressures we are concerned

about it is sufficient to assume an ideal gas.

The refractive properties of neon are not quite as well known, but can still be found from standard references. In this work we use a value of 1.000065934 [86] (632.8 nm)[1].

As an example, calculating the weighted average refractive index around the UG-3 cavity with a gas mixture of 0.2 mbar Ne, 3.3 mbar He gives a refractive index of $1+6\times10^{-8}$. Over the length of the UG-3 laser, this constitutes just 7λ. The uncertainties in the refractive index are in no way an impediment to wavelength level perimeter measurement.

3.10. Summary and choice of gas parameters

Given that the first operational helium-neon laser was reported in 1961 [89], it is remarkable how much remains to be learned about the system. In choosing the gas mixture parameters and calculating the gain curves we still rely on phenomenological models from literature. These were often determined many years ago, typically before widespread commercialisation of the helium-neon laser in the 1970's. These models do not always depend on parameters which we know they should be influenced by. In particular, very little attention has been paid to the gas temperature, which can vary significantly under typical operating conditions.

While quite some effort was made to determine the dependence of the saturation intensity on gas parameters, the results to date are barely any more definitive than historical results. There is still a good argument for revisiting these measurements.

Following this review, it is clear that there is no single factor determining the optimal gas mixture for a ring laser gyro. There are a number of competing effects which must be balanced to choose the best gas mixture for a given instrument.

High gas pressures (up to around 10 mbar) have some advantages:

- The broader gain curve allows for greater detuning without the possibility of a 'mode jump' as the cavity geometry is perturbed. This is particularly advantageous for heterolithic ring lasers.
- The dispersive pulling is reduced with a wider gain curve.

[1] Note that several web based references give incorrect values for this.

Although lower gas pressures (down to around 3 mbar) also have some advantages:

- The excitation is more efficient.
- The gas will last longer as contaminants build up, allowing longer measurement periods.
- Because of the more efficient excitation, the gas temperature is lower, this in turn minimises any gas flow or temperature dependent null shift effects.

The ratio of helium to neon appears to be not particularly critical. The excitation efficiency (Equation 3.2) and collisional broadening show only minor dependence on the He:Ne ratio between about 5:1 and 10:1. Generally we recommend using Equation 3.2 to determine the optimal ratio. The slight improvement in excitation efficiency should lead to longer gas life and lower gas temperature.

In this chapter, we have mostly restricted the discussion to neon with the natural isotope ratio. This has traditionally be used in the UG lasers and has been used in all the experiments reported here. It is also possible to use an equal isotope mixture. One advantage of this is that the wider overall gain curve allows a greater de-tuning range as the cavity geometry changes. One disadvantage in a laser with the volume of UG-2/UG-3 is that the cost of the gas is not insignificant, particularly if experiments are done which require multiple gas fills.

4. Scale Factor Correction

The approach taken when constructing the ring lasers G and C-II was to design a monolithic device from Zerodur™ for which it is known that the scale factor will remain constant to high precision. The Sagnac frequency can then be used directly as a measurement of rotation rate. The disadvantage with this approach is the excessive cost[1], particularly for devices with side lengths significantly exceeding 4 m. Practical ultra large devices must employ conventional materials following the 'heterolithic' design as has been demonstrated with the UG series of lasers.

As we have seen in earlier sections, changes in geometry are the primary source of errors in heterolithic ring laser gyros[2]. The goal of this chapter is to describe a series of experiments that have been done to validate several ideas for obtaining measurements of area and perimeter; and then to discuss their applicability to a next-generation ring laser gyro.

Area and perimeter measurements could be used to provide *scale-factor correction* of the raw Sagnac frequency measurements from a heterolithic ring laser. Employing effective scale-factor correction as part of routine operation could lead to an unprecedented improvement in performance. Such a device could potentially have performance similar to a hypothetical monolithic device of the same size.

4.1. Laser cavity deformations

Full correction for all possible deformations that the beam path could be subject to would require monitoring the movements of each reflected spot and thus each mirror in five degrees of freedom[3]. Relative displacements of the mirrors parallel to the laser

[1] Of the order 1M USD/m^2 with machining.

[2] Assuming measurement periods from seconds to several days.

[3] Rotation about the normal to the mirror surface would not be expected to change the scale factor, though might introduce changes to the Sagnac frequency via variations in backscatter pulling.

plane, and corresponding changes in the side lengths occur because the mirrors are coupled in this plane through the rock on which the instrument is sited. The rock is subject to strains and localised disturbances.

Displacements perpendicular to the laser plane are expected to occur through daily atmospheric pressure variations. To first order, these are expected to push vertically on each corner of the laser and thus not change the Sagnac frequency. However, this neglects localised variations in the elasticity of the rock and complicated interactions such as deformations of the cavern through the *cavity effect* (see [19]). Atmospheric pressure changes have been known to cause measurable tilts of the order of $20\,\mathrm{nrad\,hPa^{-1}}$ in the Cashmere cavern [19].

In this chapter we restrict the discussion primarily to *in-plane* displacements of the mirrors and changes in the angles of incidence which cause changes in the geometry of the laser. Note that this does not represent an assumption that the vertical displacements *of the beams* and the resulting changes in the overall laser plane has negligible effect. Correction for these changes has been previously demonstrated by Pritsch *et al.* [90] with UG-2 using CCD cameras to measure beam positions. They have shown that the resulting corrections can yield a dramatic improvement in the medium term[1] Sagnac frequency stability.

Let us first consider a deformation of the laser cavity geometry following displacement of the corners. Assuming the rock on which the laser is sited is homogeneous, the deformation can then be described by a linear combination of four *uniform strain fields*. These are represented in terms of a displacement vector $\mathbf{u}$. The partial derivatives $\frac{\partial \mathbf{u}_i}{\partial \mathbf{x}_j}$ which form the displacement gradient tensor associated with the possible deformations of a rectangular cavity are shown in Table 4.1. These four deformations can be considered a basis set from which an arbitrary deformation can be described.

In most cases it is reasonable to hypothesise that these deformations are excited by independent physical processes, and that there is a level of cross-coupling between them. In the case of high frequency deformations caused by seismic events, the assumption of a uniform strain field is not really reasonable. Here the processes exciting the individual components of the deformation are certainly not independent.

[1] Hours to days.

However, if the distance to the epicentre or the seismic wavelength is very large compared to the laser dimensions (almost always the case) then it is reasonable to assume that $\mathbf{u}$ is sufficiently smooth that the first spatial derivatives can be taken as constant over the dimensions of the laser.

Consider a rectangular ring laser like UG-3 which has beam combiners and Sagnac detectors at all corners. We can think of the standing wave pattern as remaining fixed in inertial space with the detectors rotating around it. Should the length of one of the arms change, there will be a corresponding change in the relative phase of the Sagnac signal observed at the adjacent corners. By tracking these relative phases we can identify relative length changes in the arms.

- The observation of pure rotational deformation is accessible through an equal shift of the Sagnac phases at all corners or equivalently by a change in the average Sagnac frequency while the rotation is happening.
- The observation of dilation is possible by observing a change in the perimeter through a change in the average wavelength of the co-rotating beams.
- The observation of rectangular shear is possible by observing symmetric changes in the lengths of opposing arms.

The observation of diagonal shear is unfortunately not possible for a rectangular laser given arm lengths alone. Consider the lengths of the horizontal arms following this deformation as shown in Table 4.1. Under this deformation, we would observe a change in length in the horizontal x direction of

$$dL = \sqrt{\left(dx + \frac{\partial \mathbf{u}_x}{\partial x} dx\right)^2 + \left(\frac{\partial \mathbf{u}_y}{\partial x} dx\right)^2}. \tag{4.1}$$

This change in length is fundamentally indiscernible from that associated with a dilation, unless the angles of incidence can be measured[1].

In the absence of measurements of the angles of incidence and assuming negligible change in perimeter[2], we can obtain a good approximation to the area using Brahmagupta's formula. This gives the area A of a cyclic quadrilateral (a simple quadrilateral inscribed in a circle) given side lengths of the north L_N, east L_E, south

[1] Section 4.3 details a possible method for determining this parameter.

[2] Limits can be placed on the perimeter changes by the frequency of mode jumps.

L_S and west L_W arms (see Figure 4.1)

$$A = \sqrt{(s - L_N)(s - L_E)(s - L_S)(s - L_W)} \tag{4.2}$$

where s is the semi-perimeter

$$s = \frac{L_N + L_E + L_S + L_W}{2}. \tag{4.3}$$

4.1.1. Aside: deformations of a triangular laser cavity

If we repeat the cavity deformations shown in Table 4.1 for a triangular laser it is interesting to note that we can uniquely determine all deformation components if given the relative phase shifts at each corner and the operating wavelength.

The area of a triangle, regardless of deformation can always be determined from just the side lengths using Heron's formula [91], and thus if we assume a triangular laser built perpendicular to the rotation axis (so that changes in tilt have negligible effect) then the Sagnac frequency given side lengths a, b and c is always

$$f_s = \frac{\sqrt{(a+b+c)(-a+b+c)(a-b+c)(a+b-c)}}{a+b+c} \frac{\Omega}{\lambda}, \tag{4.4}$$

and unlike the rectangular laser, all of the geometric parameters can be determined to interferometric accuracy. This has not been considered historically; the rectangular geometry has traditionally been used to maximise scale factor within the physical constraints of the available space.

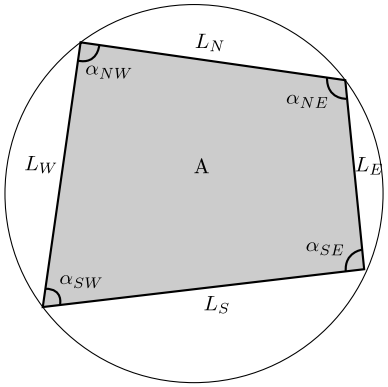


Figure 4.1. Illustration of the notation conventions used in this thesis for describing the geometric measurements associated with scale factor changes.

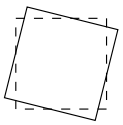
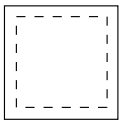
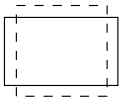
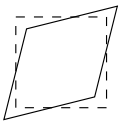
Deformation	Illustration	Strain Field	Discernible?
Rotation		$\frac{\partial \mathbf{u}_x}{\partial y} - \frac{\partial \mathbf{u}_y}{\partial x}$	✓
Dilation		$\frac{\partial \mathbf{u}_x}{\partial x} + \frac{\partial \mathbf{u}_y}{\partial y}$	✓
Rectangular shear		$\frac{\partial \mathbf{u}_x}{\partial x} - \frac{\partial \mathbf{u}_y}{\partial y}$	✓
Diagonal shear		$\frac{\partial \mathbf{u}_x}{\partial y} + \frac{\partial \mathbf{u}_y}{\partial x}$	✗

Table 4.1. Two-dimensional deformation components for a rectangular ring laser cavity.

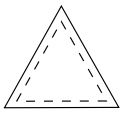
Deformation	Illustration	Strain Field	Discernible?
Rotation		$\frac{\partial \mathbf{u}_x}{\partial y} - \frac{\partial \mathbf{u}_y}{\partial x}$	✓
Dilation		$\frac{\partial \mathbf{u}_x}{\partial x} + \frac{\partial \mathbf{u}_y}{\partial y}$	✓
Rectangular stretch		$\frac{\partial \mathbf{u}_x}{\partial x} - \frac{\partial \mathbf{u}_y}{\partial y}$	✓
Shear		$\frac{\partial \mathbf{u}_x}{\partial y} + \frac{\partial \mathbf{u}_y}{\partial x}$	✓

Table 4.2. Two-dimensional deformation components for a triangular ring laser cavity.

4.2. All-corner phase tracking

Phase tracking (as detailed in Section 6.6.2) has been applied to every corner of UG-3 at an acquisition rate of 5 Hz as part of routine data acquisition for the operational history of this laser configuration.

The accumulated phase measurements have been analysed to determine the side lengths for several runs in which the laser was minimally disturbed.

An increase in phase relative to the background Sagnac signal from earth rotation indicates a rotation with the earth. In our case the background rotation is clockwise when viewed from above so for example an increase in phase of 0.1 cycles relative to a stable reference on the north-east corner will indicate an increase in length of the north arm by $0.1\frac{\lambda}{2}$ (recall that the standing wave has periodicity $\lambda/2$).

The procedure for obtaining the length measurements is as follows:

1. A linear least-squares fit was applied to each accumulated phase record. The average coefficients from the four corners were found and these used to subtract a long term drift from each record. This is necessary because the reference generator will never be set to exactly the nominal Sagnac frequency.

2. The residuals of the phase records ϕ were differentiated with respect to time t to obtain $\frac{\mathrm{d}\phi}{\mathrm{d}t}$.

3. Working clockwise around the laser, the difference in phase $\Delta\phi$ over each arm relative to the preceding corner was found as a function of time.

4. The changes in length were found by cumulative summation and these added to the nominal length L_0, i.e.

$$L(t) = \left(\sum_{i=0}^{i=t/\mathrm{d}t} \frac{1}{2}\lambda\,\Delta\phi(i) \right) - \bar{L} + L_0 \tag{4.5}$$

   The nominal length is known from a survey (see Figure 2.1), and this analysis assumes the lengths converge to these values over the long term. $\bar{L}$ is the long term average of the length changes.

5. For those measurements where the long or medium-term behaviour was of interest, the measurements were smoothed using a moving average weighted with

a Hanning window function of width 1000 s. The signals were prepared for smoothing by introducing reflected copies of the end portions (each one window length) at each end. This allows the original data length to be retained and minimises any transient effects at the ends.

4.2.1. Strain measurements

On all time scales from a few seconds to several days the extension of parallel arms remains very strongly correlated. This observation provides good evidence for the validity of the uniform strain/affine deformation hypothesis. On occasions where the extensions of parallel arms is not well correlated, it can be concluded that a localised disturbance of the laser occurred.

Over short time periods the strain measurements are dominated by microseismic disturbances (Figure 4.2) with frequency of 100 mHz to 200 mHz and strain amplitude of the order 1×10^{-10}. The statistical properties (representative frequency and amplitude) of the microseismic motions varies considerably from day to day.

Over long time periods the strain measurements are dominated by a ∼12 hour semi-periodic motion (Figure 4.3). This is caused by deformations of the earth due to tidal forces from the sun and moon, usually termed *solid earth tides*. These are further discussed in Section 5.3. A program *pysolid* (documented in Section A.3 and Section A.3.1) is available to simplify calculations of their effects. This was used here to compare the measured strains to theoretical predictions.

The east-west strains show a clear agreement with the direct theoretical prediction, particularly in the locations of the turning points. The RMS variations measured in the north and south arms are factors of 2.5 and 2.65 (respectively) larger than predicted by the model.

In reviewing strain measurement instrumentation and techniques, King and Bilham [92] have noted that the properties of the rock on which the instrument is sited can have a large effect on the measured strain amplitudes. They proposed a model[1] in which Young's modulus E varies periodically (RMS variation E_r) from the nominal

[1] The geophysics involved in this model is complex and outside the scope of this thesis.

value E_0 with respect to distance x along the strain vector,

$$\frac{1}{E} = \frac{1}{E_0} + \frac{\sqrt{2}}{E_r} \cos \frac{2\pi\, x}{\Lambda}. \tag{4.6}$$

They note that while this one-dimensional model might appear at first to be so idealised as to be of limited application, it does cover most of the likely (and some unlikely) inhomogeneity geometries that would be expected geologically provided that there are no large fissures or cavities and provided that an error of up to 30% is allowed for.

Using this model it is possible to relate our measured fractional error in the variations from theory to the wavelength Λ and peak fractional difference in Young's modulus E_0/E_r for the periodic variations using the expression [92],

$$\Lambda = \frac{\text{RMS(observed)}}{\text{RMS(theory)}}\, \pi L \left(\frac{E_0}{E_r}\right)^{-1}. \tag{4.7}$$

King and Bilham [92] note that the variations in Young's modulus might typically be 30% or more. Figure 4.4 plots the wavelength of periodic variations in Young modulus for a range of reasonable values. It appears from this that variations on the scale of 30 m to 50 m could explain our results.

The north-south strains are much more poorly predicted by tidal models. These are best modelled as phase-shifted versions of the east-west strains, suggesting that there is considerable cross coupling taking place. King and Bilham note that inhomogeneities such as fissures can cause large cross coupling and peak shifting effects in the measured strain signals.

One particularly interesting type of inhomogeneity (first outlined by Panck [93]) is the presence of a tunnel in which the strain measurement is being made. Panck notes that for an ideal tunnel, there will be no effect on the measurement of strain along the axis of the tunnel. However, orthogonal to the axis of the (ideal) tunnel, they show that the measured strain amplitudes will be amplified precisely by a factor 3.

The north and south chambers of the Cashmere cavern can be roughly approximated by two tunnels, and the strains measured parallel to the tunnels show good agreement with the theory. While the orthogonal strain waveforms do not agree with theory, the RMS variations from the mean generally do, indicating that this amplification factor does not need to be applied.

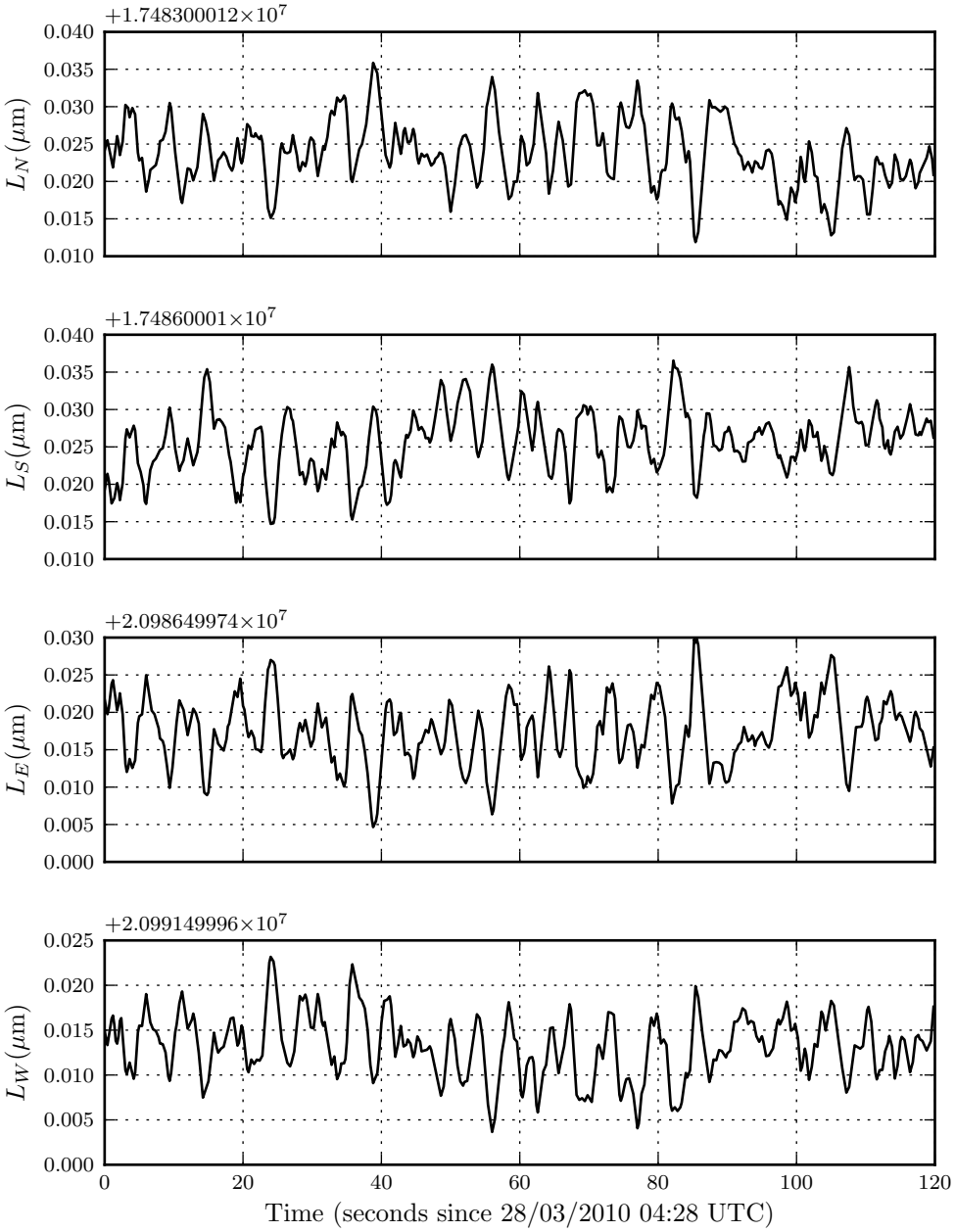


Figure 4.2. Observed short term stains due to microseismic activity. Note the strong correlation of the parallel arms.

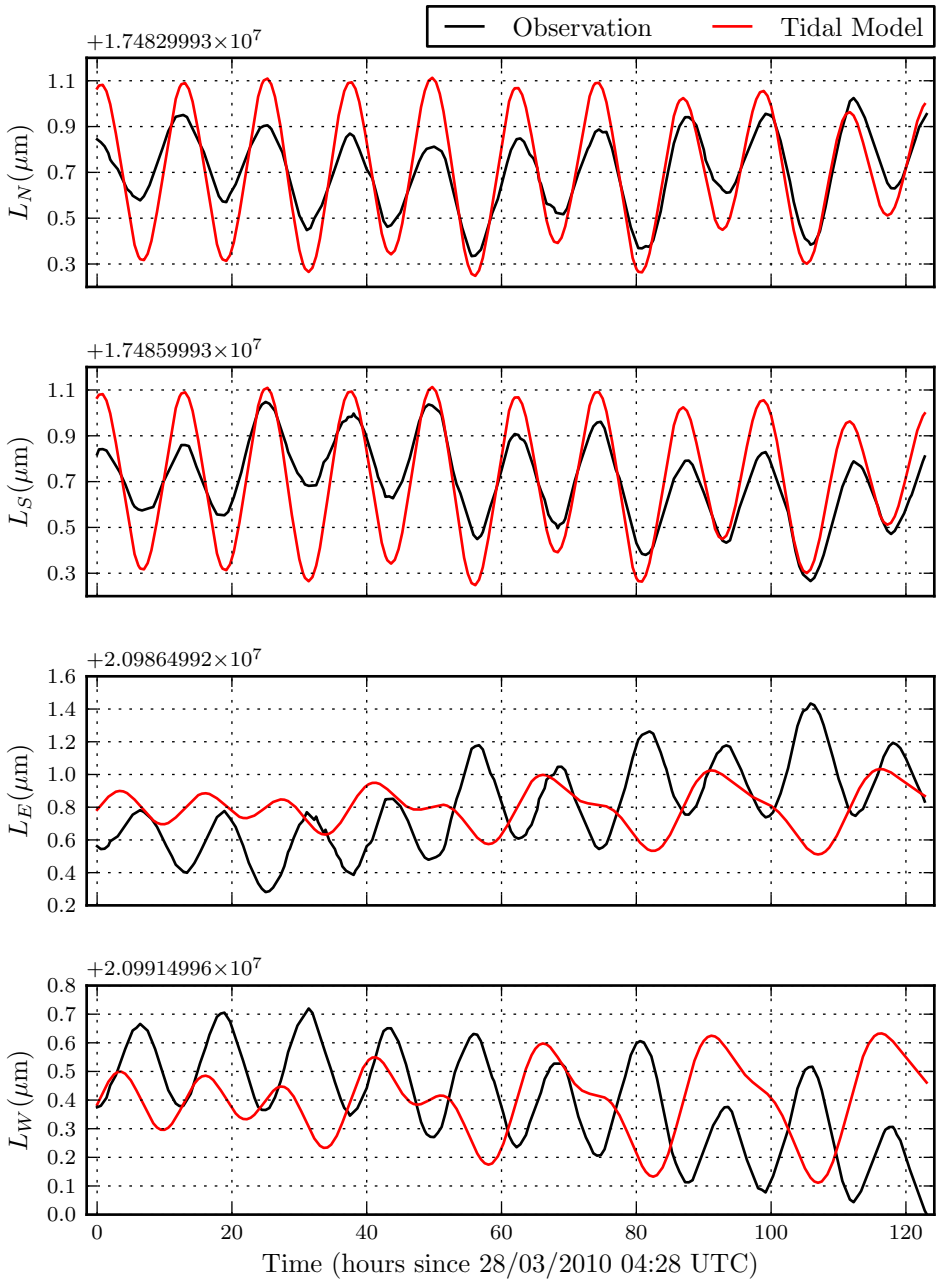


Figure 4.3. Observed and theoretical arm lengths over 5 days on UG-3. Note that changes in the north and south arm lengths are caused by strains in the east-west direction and vice-versa. Note that while this data set shows very sinusoidal-looking behaviour, this is not always the case and the tidal strain signals are best described as *semi-periodic*.

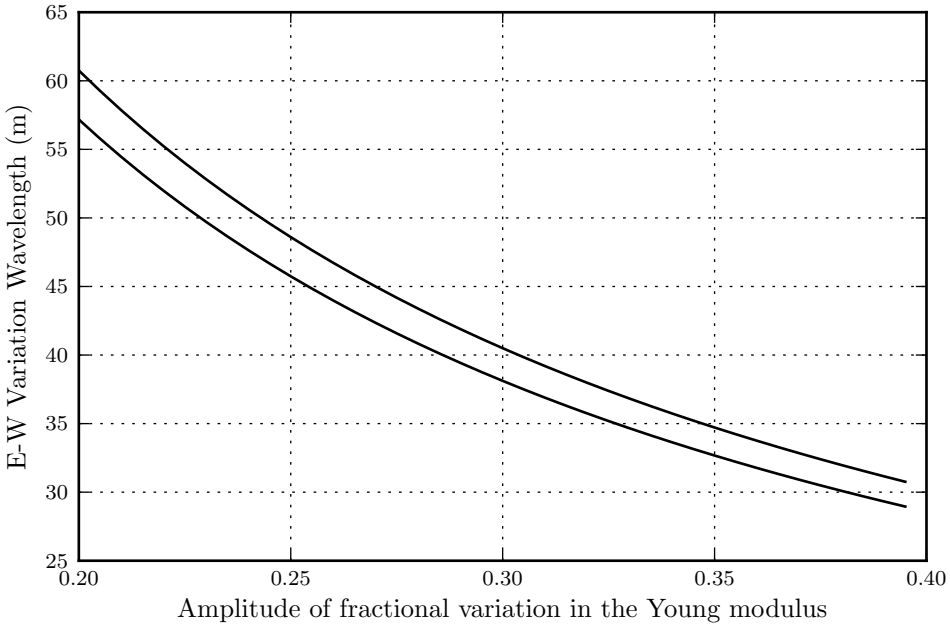


Figure 4.4. Possible parameters of a east-west periodic variation in the Young modulus which could explain our observed differences in theoretical length variations of the north and south arms. The lower and upper bounds were calculated from the north and south arms respectively.

4.2.2. Sagnac frequency correction

Using Equation 4.2, the geometric part of the scale factor, A/P has been calculated for the data set shown in Figure 4.3. A corrected Sagnac frequency time series f_s' has been calculated given the nominal area A_0 and nominal perimeter P_0,

$$f_s' = f_s \frac{AP_0}{A_0 P}. \tag{4.8}$$

This is shown in Figure 4.5. The maximum correction calculated this way is 5 µHz. Note that while the measured Sagnac signal does appear to have components at similar frequency to the tidal strains, these are most probably caused by tidally induced tilts which have a much larger effect (±150 µHz) and possibly also atmospheric pressure cycles which also have a ∼12 hour component.

4.2.3. Summary of all-corner phase tracking

Overall, all-corner phase tracking is easy to implement and provides useful strain information. It also allows an average Sagnac signal to be computed that minimises the effects of localised disturbances of the instrument or environment.

As with other measurements of strain, the quality of the site is critical because inhomogeneities can cause large changes in the measured strains. For UG-3, this is particularly true in the north-south direction, though not for the east-west direction, where the measured strains agree well with theoretical predictions computed from a model of tidal earth strain. Having high-rate strain measurements available during seismic events potentially adds useful information for understanding these events.

While not a primary research interest, good measurements of tidal strain are interesting from a geophysical perspective. This is likely to be the case more so in the future, following the recent work of Thomas *et al.* [94], in which a correlation was found between extremely small, tidally induced shear stress parallel to the San Andreas fault and non-volcanic tremors in the region.

Using relative side-length measurements to derive a scale factor correction is easily done, though there is little point in applying it on a routine basis to the measurements from UG-3 at present. This is not because it doesn't work well, rather it is because the corrections it offers are small and presently dominated by other noise sources.

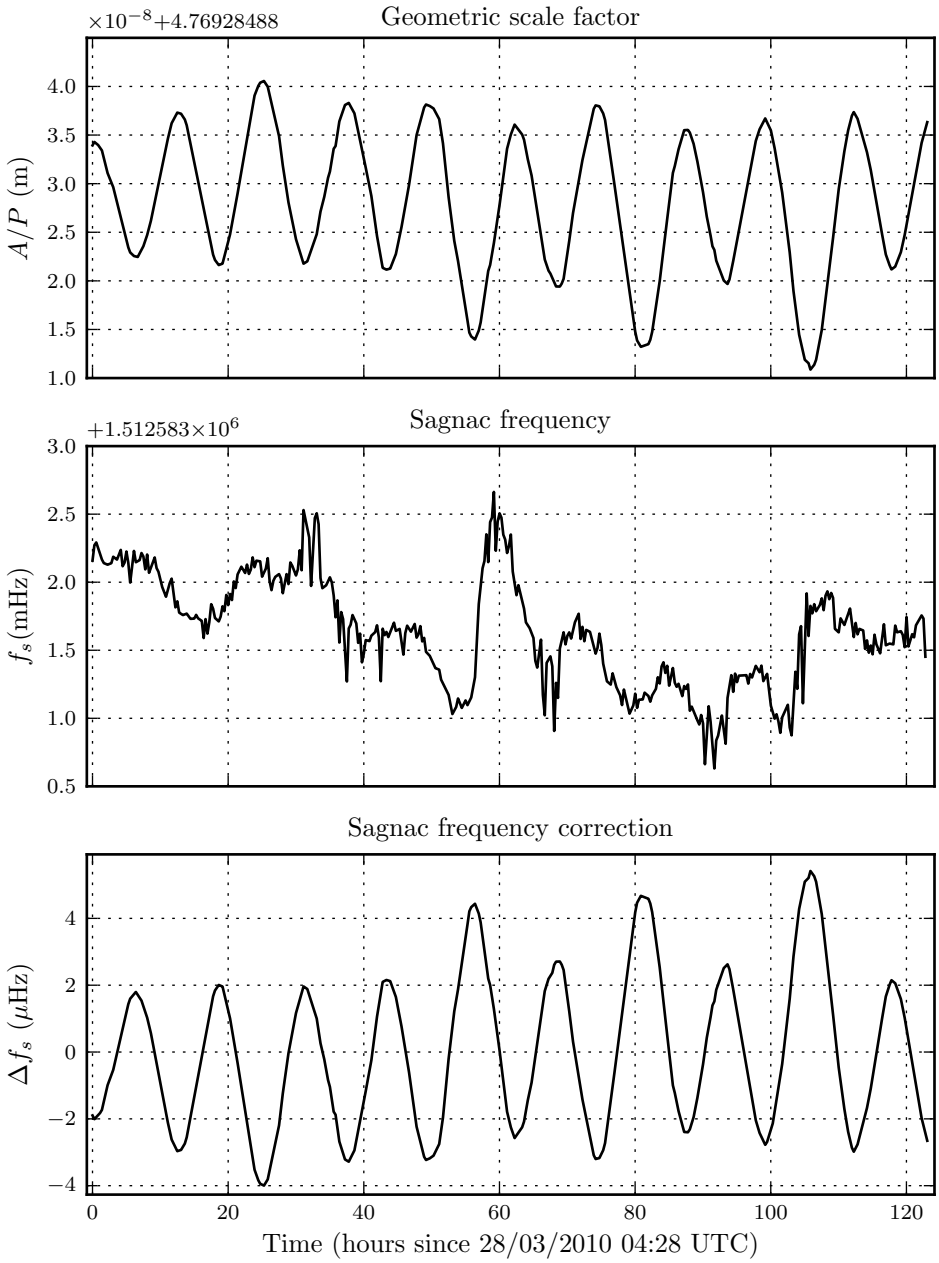


Figure 4.5. Measured Sagnac frequency over 5 days on UG-3 with a correction calculated by Brahmagupta's formula. Note that the corrections are always less than 1% level of the variations due to other factors such as tilt, atmospheric pressure.

However, if these could be eliminated in a next generation ring laser gyro, scale factor correction by all-corner phase tracking could well be usefully applied.

Following these results and the work of Pritsch *et al.* [90], it is clear that any next-generation ultra large ring laser should be constructed so that the laser plane is mathematically well defined. Such a laser would require a feedback system to control the tilt and rotation of the mirrors to ensure the beams are always in the centre or 'clear aperture' of the laser mirrors. This arrangement is sometimes termed the 'virtual monolith'.

4.3. Sagnac contrast

We define the *Sagnac contrast* c_s as the ratio of the amplitude of the Sagnac signal to the mean value. In this section we consider the usefulness of Sagnac contrast as a measure of beam combiner alignment, and thus as a way to determine the angles of incidence on the mirrors.

Consider the Gaussian intensity distributions I_a and I_b due to two TEM_{00} beams of width w separated by s,

$$I_a(r,s) = \frac{1}{(w^2/2)} \exp\left(-\frac{2(r-(s/2))^2}{w^2}\right) \tag{4.9}$$

and

$$I_b(r,s) = \frac{1}{(w^2/2)} \exp\left(-\frac{2(r+(s/2))^2}{w^2}\right) \tag{4.10}$$

The contrast ratio can be found from the ratio of the intersecting region (which is responsible for the beat frequency) to the total amplitude,

$$c_\mathrm{s}(s) = \frac{\int \min\left(\sqrt{I_a(r,s)}, \sqrt{I_b(r,s)}\right)\, \mathrm{d}r}{\int \frac{1}{2}\left(\sqrt{I_a(r,s)} + \sqrt{I_b(r,s)}\right)\, \mathrm{d}r}. \tag{4.11}$$

This is illustrated in Figure 4.8. Equation 4.11 can be evaluated for $w = 2.1\,\mathrm{mm}$ (UG-3) (Figure 4.9). From this we see that for beam separations approaching 1 mm, the contrast ratio is very close to linear with respect to separation. In our case the separations will always be much less than this, so

$$s = 3.762(1 - c_s) \tag{4.12}$$

gives a very good approximation of the separation given the contrast ratio.

In addition to beam combiner alignment, Sagnac contrast can also be reduced by background light, wavefront distortions, detector 'dark current', detector bandwidth limitations, obstruction causing the beam to not completely fall on the active area of the detector and any differences in size or divergence of the individual beams.

To first order, it is possible to eliminate changes in Sagnac contrast due to movements of the beam across the mirror by careful design of the beam combining optics. We have standardised on a '2-reflection' arrangement for the counter-clockwise beams for all the work with UG-3. Figure 4.6 compares this arrangement to a simpler '1-reflection' arrangement and illustrates how movement of the reflection point on the mirror will cause the individual beams to track each other and thus not change the Sagnac contrast. This arrangement also prevents beam movements generating phase changes in the combined signal.

The '2-reflection' arrangement is most elegantly implemented with a pentaprism. In practice we find that with careful alignment of this arrangement, it is possible to achieve very near to an ideal (unity) Sagnac contrast.

If L is the nominal distance from the laser mirror to the detector (usually $\sim$380 mm in UG-3) then the angle of incidence is simply

$$\alpha = \alpha_0 \pm \sin^{-1} s/L \tag{4.13}$$

where $\alpha_0 \approx \frac{\pi}{2}$, the nominal angle of incidence (when the beam combiner is ideally aligned). This can be found from survey data.

The main problem with this technique is that because we are only sensitive to changes in angle, we have no way of determining if the angle has increased or decreased relative to the nominal value.

Another problem that occurs when implementing this on UG-3 is that movements (in particular, rotations) of the corner box and mirror relative to the beam combiner optics (which are mounted separately) can cause changes in the contrast ratio. This is also a problem when implementing any beam position detection scheme, for example as part of a 'virtual monolith' arrangement. For this reason it is recommended that in future ring lasers, the laser mirror and optical breadboard supporting the combining optics and position detection system should be rigidly mounted together.

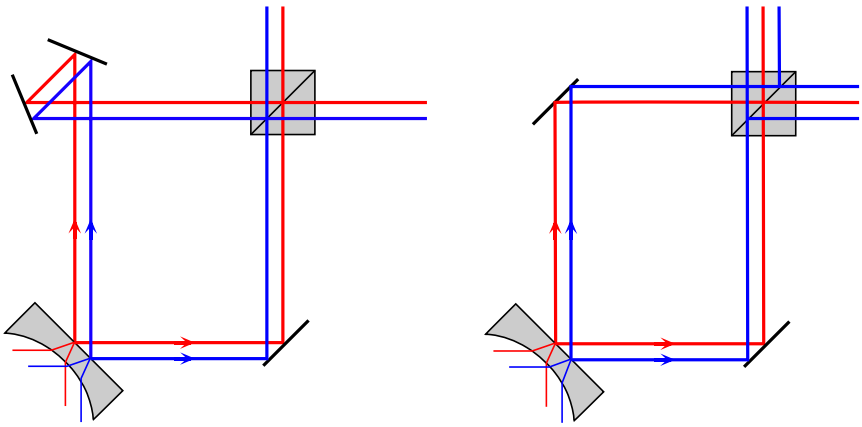


Figure 4.6. Beam translations compared in a 2-reflection beam combiner arrangement (left) and a 1-reflection arrangement (right).

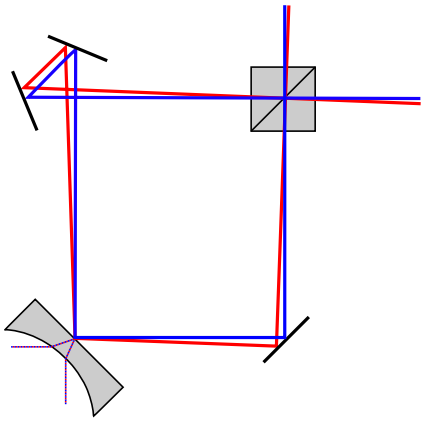


Figure 4.7. Angle of incidence shift in a 2-reflection beam combiner design.

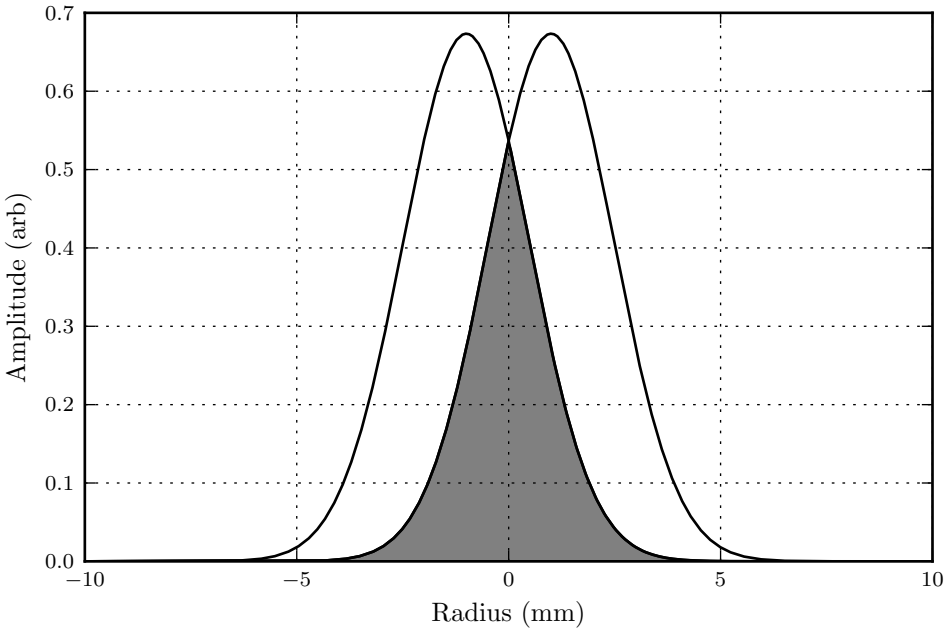


Figure 4.8. Illustration of the fraction of overlapping beams that contributes the AC component of the Sagnac beat signal.

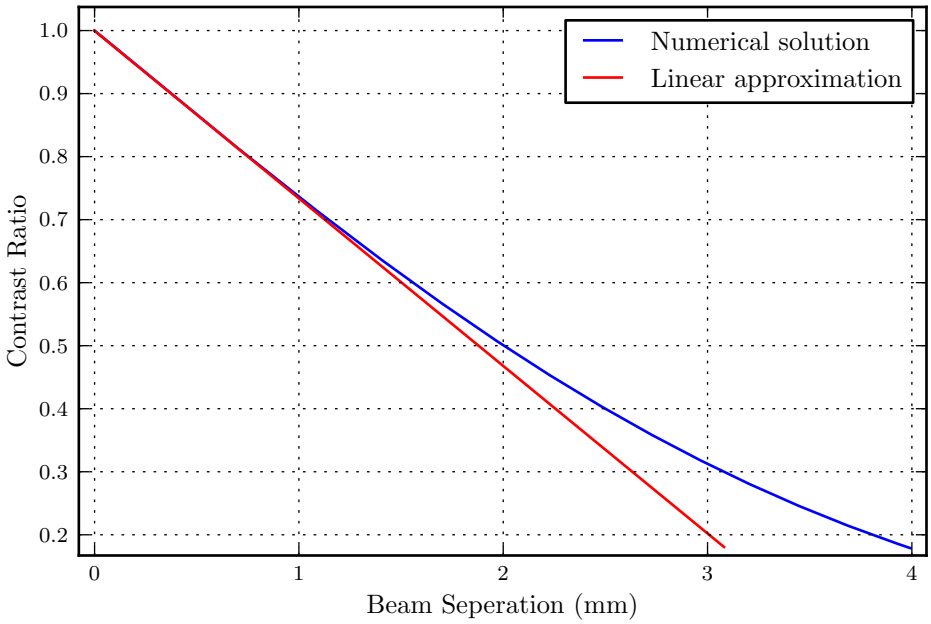


Figure 4.9. Relationship between contrast ratio and beam separation.

Integration time	Average standard deviation	Standard deviation of average
1 s	4.8 µrad	2.6 µrad
10 s	3.6 µrad	1.9 µrad
100 s	3.3 µrad	1.8 µrad
1000 s	3.15 µrad	1.6 µrad

Table 4.3. Uncertainty in angle of incidence measured by Sagnac contrast.

4.3.1. Sagnac contrast measurements

Contrast ratios have been acquired at a rate of 5 Hz as part of the routine data acquisition with UG-3. Angles of deviation from the long term average have been calculated for each corner and a typical data set is shown in Figure 4.10.

We observe no clear relationships or trends in the time series for the angles of incidence at any of the corners over periods from seconds to days. The time series were analysed using the method of autocovariance for periods of up to 30 h and no repeating patterns were found.

If we are indeed measuring angles of incidence, we should expect the angles at all corners to add up to 2π. Given this, the *variations in the average* angular deviation should be smaller than the *average of the variations* in the individual angular deviations. Given a 15 h data set selected for minimal disturbance, we observe that this is indeed the case (Table 4.3). The latter is always around half of the former regardless of averaging time. The latter provides a measurement of the uncertainty in the former, i.e. over an averaging time of 1000 s we are able to resolve changes in angle to a precision of 3.15 ± 1.60 µrad.

The typical daily variation in the angles of incidence measured this way is ∼10 µrad. This corresponds to a beam movement at 20 m of 0.2 ± 0.1 mm. It is interesting to compare this to the beam movements observed by Pritsch *et al.* [90] on UG-2 using CCD cameras. The extent of typical movements observed was of the order 0.1 mm. Note that this method could be implemented on UG-3 using the spare combined beam, although it would not respond to angle of incidence changes and may be difficult to implement due to beating effects between the CCD scan rate and the Sagnac modulation inevitably present on the combined beam.

Brief, large disturbances which occur on all corners have been observed. These are most probably due to mechanical shocks and changes in background light levels.

Disturbances characterised by a rapid onset followed by a slow decay (up to 3 hours) back to the nominal value have been observed. These are not always observed on all corners simultaneously. This type of disturbance is most probably caused by thermal expansion in the lever arms following the presence of people near respective corner.

4.3.2. Summary of Sagnac contrast monitoring

Overall, we have demonstrated the concept of measuring angles of incidence by Sagnac contrast and shown that changes in angles of incidence of a few µrad can be resolved. Sagnac contrast measurement is a useful method for determining the overall statistical behaviour of the angles of incidence, but it is not a useful method for providing a correction for scale factor perturbations. This is because the direction of the change in angle can not be resolved and because the measurement is very vulnerable to disturbances.

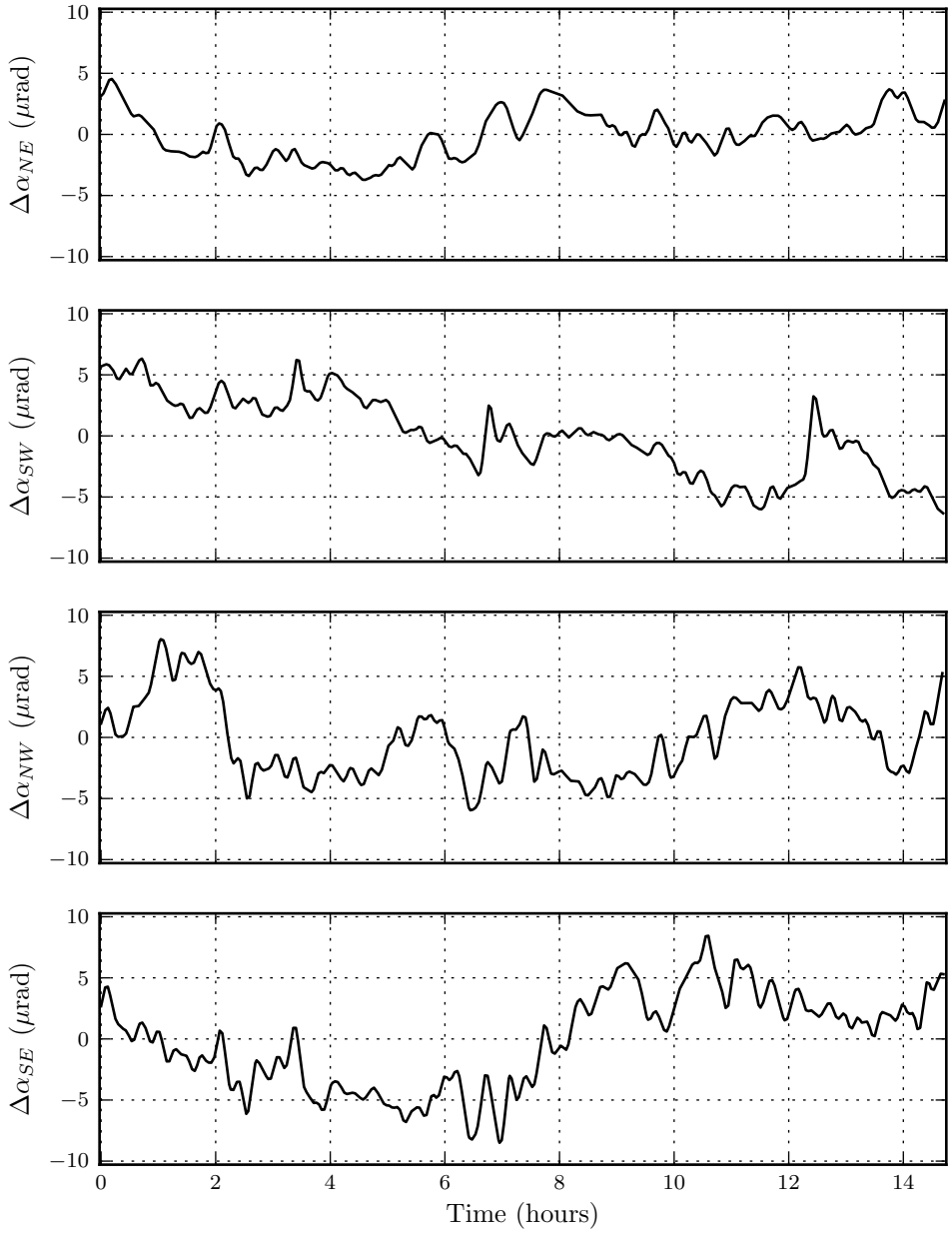


Figure 4.10. Long term (averaged with a 1000 s Hanning window) behaviour in angles of incidence measured by contrast ratio showing no obvious relationships or trends.

4.4. Multi-mode operation

Multi-mode configurations are interesting because they could potentially provide both a Sagnac signal and a way to measure the FSR, which could in turn be used to correct the measured Sagnac frequency for perimeter changes. Additionally, they could also allow for higher operating power (no longer limited by the multi-mode threshold) which could improve the overall signal to noise ratio and operational convenience. This section outlines potential techniques to exploit multi-mode configurations for perimeter correction.

4.4.1. Performance under mode-coupling

Before considering any schemes for exploiting operation in the mode-coupled regime (see Section 3.4.4), it is first necessary to establish whether any difference in performance exists.

The performance of the UG-3 laser has been characterised by calculating the Allan variance [29] from Sagnac frequency data obtained using the phase tracking method (Section 6.6.2) for 3 runs in the regular single mode regime and 3 runs in different mode-coupled configurations. Because of the variability in each run, the data sets have been re-sampled using cubic spline interpolation and an averaged data set presented in each case with 1σ error bars indicating the variability. This is shown in Figure 4.11.

In Figure 4.11 we notice a slight decrease in ability of the laser to respond to fast micro-seismic signals when mode-coupled though overall we observe no statistically significant difference in performance.

4.4.2. Sagnac frequency under mode-coupling

It is also important to determine if there are any systematic shifts in the Sagnac frequency when operating in the mode-coupled regime.

Averaging the Sagnac frequencies recorded on UG-3 from the above mentioned runs we find that the Sagnac frequency is systematically 7.5 ± 5.0 mHz lower than normal. A larger and less predictable frequency shift of as much as 1 Hz higher or lower has also been observed under similar circumstances on the G-0 laser.

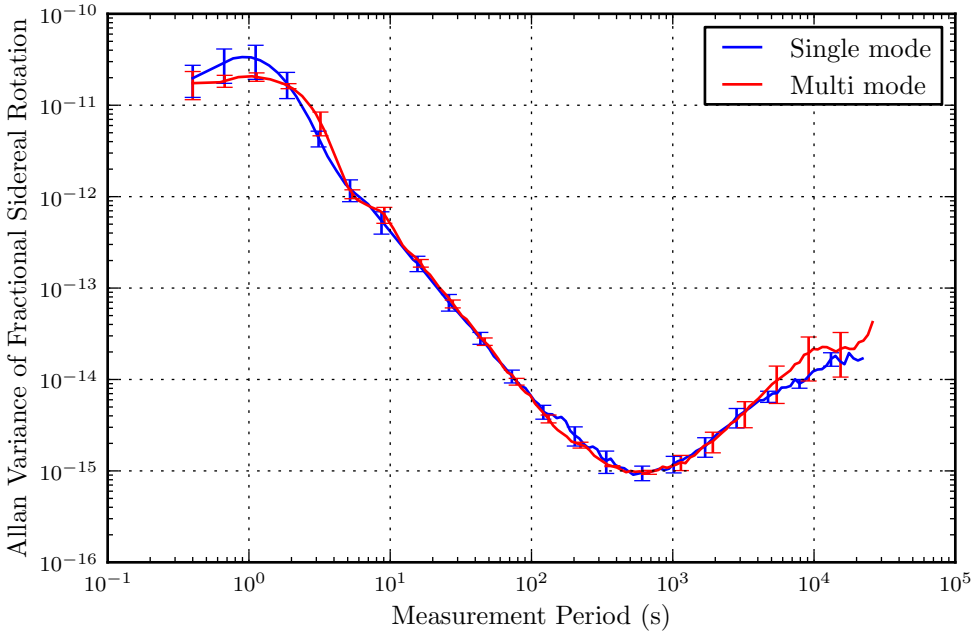


Figure 4.11. Mode-coupled performance comparison of UG-3 by Allan Variance.

Several potential mechanisms might cause a change in the Sagnac frequency.

It is not yet understood how the wavelengths of the various operating modes should be weighted to determine an effective operating wavelength λ. However if we calculate the change in Sagnac frequency associated with the most extreme mode (8×FSR) we see this can only account for a Sagnac shift of 101 µHz (UG-3 laser). This can not explain the measured discrepancy.

When operating in the mode-coupled regime the gain medium is hotter due to the higher operating power. As outlined in Section 3.6, there is a small temperature related shift in the optical frequency at the peak of the composite gain curve. This has been calculated for the natural neon isotope mix used in UG-3 at the time of measurement (see Section 3.3). The optical frequency shift was found to be 100 kHz K^{-1}. Accounting for a 7.5 mHz Sagnac shift would thus require an impossibly large temperature shift of 23.5×10^3 K. So this effect can not explain the measured discrepancy.

As has been detailed in Section 3.8, slight asymmetries in the electrode configuration could result in a net flow of gas which in turn could differentially bias the optical

frequencies of the counter-rotating beams and introduce an offset in the Sagnac splitting. This is simulated by adding a Doppler offset $\frac{v}{\lambda}$ to the Gaussian curve when calculating the gain curve. Solving for a frequency shift of 7.5 mHz we find a velocity of just 2.65 mm s^{-1} can account for the observed offset.

While our use of RF excitation should minimise gas flow, it seems unlikely that it could be reduced to the level of a few percent of that observed in DC discharges, especially given the sometimes striking asymmetries we observe in the plasma and the fact that the electrode configurations we use are only nominally symmetrical. Given these observations it appears that plasma flows are most likely the dominant factor contributing to the observed Sagnac frequency offset.

4.4.3. Mode-coupled configuration stability

When operating in the mode-coupled regime a Fabry-Pérot spectrum (obtained using the acquisition and processing technique detailed in Section 6.6.5) clearly shows the operating modes, separated by multiples of the FSR. Figure 4.12 shows the identified peaks corresponding to a mode configuration observed on UG-3 when operating with a strong Sagnac signal.

Figure 4.13 shows the results of an analysis of many spectra similar to Figure 4.12 recorded at a rate of 3 Hz for 70 hours. The goal of this exercise was to provide a quantitative description of the long term stability of mode-coupled configurations and to demonstrate the principle of FSR monitoring to high precision via Fabry-Pérot spectrometer.

The stability of the mode configurations has been characterised by finding the average powers of the running modes and the standard deviation of their respective powers[1]. Long term stability is a necessary condition for any technique exploiting multi-mode operation for scale factor correction. This shows that even after many hours, no long-term stable configuration is ever obtained. Note that the laser was operated under servo control to keep the total beam power constant.

The FSR was found from each spectrum by finding the frequency differences between running modes that were approximately (within 3%)[2] of a multiple of the nominal

[1]The units are volts recorded on the Fabry-Pérot photomultiplier.

[2]This is necessary because it is impossible to perfectly calibrate the Fabry-Pérot scan to avoid a

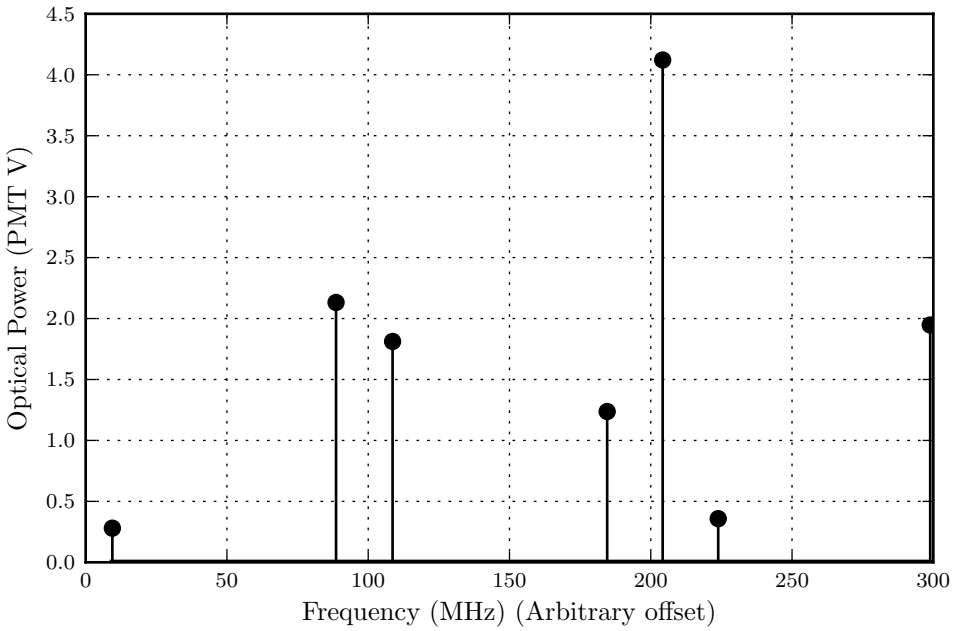


Figure 4.12. Typical mode configuration of the UG-3 laser when running in the mode-coupled regime.

FSR. The procedure used to compute the FSR from a spectrum is illustrated in Listing 4.1.

Ultimately the precision to which these frequencies can be measured is determined by the linearity of the piezo mirror movement and it is unlikely that this method would ever be competitive to optical beating techniques as a way of measuring perimeter. However we were able to demonstrate an impressive precision in the average FSR of $3.91 \pm 0.02\,\mathrm{MHz}$. Note that the uncertainty represents a fraction of 6.6×10^{-5} of the 300 MHz Fabry-Pérot scan range.

4.4.4. FSR recovery by direct beating

One method for obtaining the perimeter of UG-3 in the mode-coupled regime is to directly beat the signal from a photomultiplier with a reference oscillator in the same manner as described in Section 4.5.1. This was attempted with signals from combined and un-combined beams. Overall this method was unsuccessful. Some success was

frequency offset on 'wrap-around'.

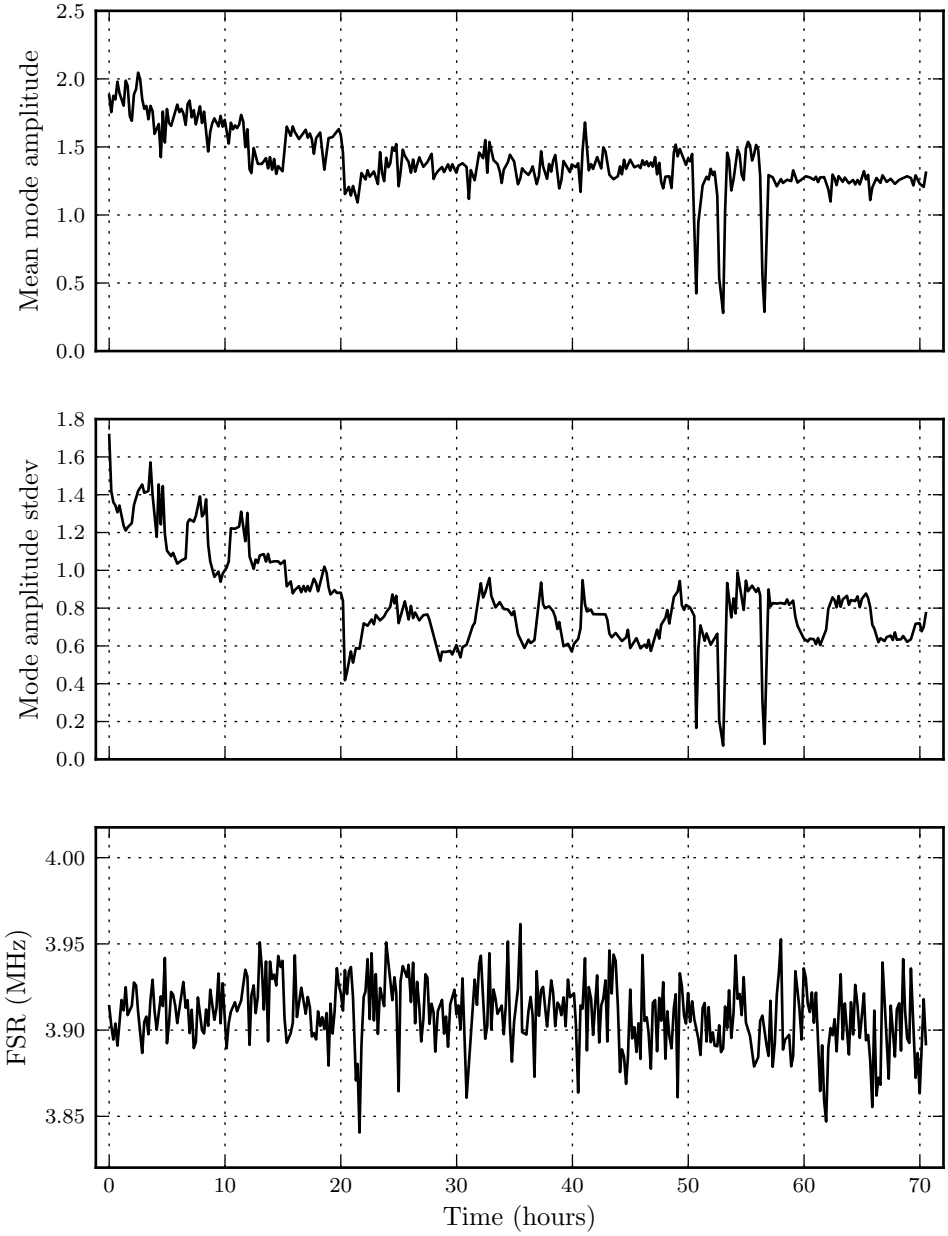


Figure 4.13. Stability of the mode configuration of the UG-3 laser running in the mode-coupled regime. Note that no tendency to long-term stability is demonstrated.

```
threshold = 0.03
fsr_multiples = []
for a in frequencies:
    for b in frequencies:
        if not a == b:
            multiple =  abs(a-b)/nominal_fsr
            remainder = abs(round(multiple) - multiple)
            if remainder < threshold:
                fsr_multiples.append(multiple)
fsr_values = (fsr_multiples - round(fsr_multiples) ) * nominal_fsr +
    nominal_fsr
```

Listing 4.1 Procedure used to calculate the FSR from a list of spectral peak `frequencies`.

noted when the operating power was set very close to the multi-mode threshold. However this configuration was not stable enough to obtain useful results. This is not surprising, if the mechanism underlying the mode-coupling phenomena operates fast enough to establish a fixed phase relationship between the RF beats in the same way as for the audio frequency beats then we would not expect to see such beats.

4.4.5. FSR recovery by Fabry-Pérot filtering

One potential technique for recovering the FSR and Sagnac signal simultaneously that was experimented with is referred to as *Fabry-Pérot filtering*. The idea here is to operate the laser well above the multi-mode threshold so that a stable 'comb' of modes is present. A combined beam is passed into a scanning Fabry-Pérot cavity with pass-band much narrower than the laser mode spacing, but wider than the Sagnac splitting (this will almost always be the case in practice). Instead of driving the mirror through a scan, a feedback system is implemented which locks on to the Sagnac resulting from the two combined modes. Meanwhile, the FSR is recovered in the usual way by an RF beat frequency technique.

A linear feedback system[1] was implemented in LabVIEW using a DAC output from a data acquisition board to control a high voltage power supply and in turn the piezo

[1]This simple approach was found to work acceptably, though a more advanced implementation could make use of the derivative signal, pre-filtered to remove the Sagnac frequency.

mirror actuator on a scanning Fabry-Pérot cavity. This technique was shown to successfully lock on to a Sagnac beat signal. However, in all cases when a stable beat frequency could be obtained, the laser was operating in the mode-coupled regime (see Section 3.4.4). This somewhat defeats the purpose of the filter cavity.

While compelling in principle, this approach was not found to work in practice with the UG-2 or UG-3 lasers. The fundamental problem with the technique is obtaining a stable 'comb' of modes without mode-coupling occurring. We note however that it is possible that other laser configurations with wider mode spacing operating at low pressure (in order to reduce the homogeneous broadening, and thus the unstable competition processes between modes) may be better suited to this technique.

4.4.6. Summary of multi-mode operation

The main practical problem with operating a UG-scale laser in the mode-locked regime is the time required to achieve stable operation. After starting the laser at a level below the multi-mode threshold, it will always initially be running with several longitudinal modes. The mechanisms of hole burning then lead to competition between these modes and if operating below the multi mode threshold, a single mode will emerge after a few minutes. The competition between multiple modes is considerably more complex and the time required to achieve even a semi-stable mode pattern can be on the order of hours. An unstable mode configuration is always accompanied by an unstable beat signal.

It is interesting to note that for a hypothetical laser larger than UG-3 or with higher Q, it is possible that the drift in perimeter over the start-up period might exceed one wavelength. In such an instrument stability could never be achieved.

4.5. Alternate-split operation

We have seen earlier (Section 3.4) that a ring laser can operate with different longitudinal mode numbers for the clockwise and counterclockwise beams (split-mode). In this configuration the laser is not useful as an instantaneous rotation rate sensor. However it is possible to use the RF beat frequency between the beams to measure their frequency difference. This will be $S\mathrm{FSR} \pm f_s$, where S is an integer (the splitting order) and the $\pm$ depends on which order the beams are arranged along the gain

curve.

Since the FSR is directly related to the perimeter

$$\mathrm{FSR} = \frac{c}{n\,P}, \tag{4.14}$$

we can use operation in split-mode to determine the perimeter, provided we know (or can estimate well) the Sagnac frequency.

Consider the case where we continually restart the laser and alternately measure $(S\,\mathrm{FSR} + f_s)$ and $(S\,\mathrm{FSR} - f_s)$ as illustrated in Figure 4.14. We then use a suitable interpolation technique to obtain continuous signals from each. From these two signals, the differential signal gives the Sagnac frequency

$$f_s = \frac{1}{2}\left((S\,\mathrm{FSR} + f_s) - (S\,\mathrm{FSR} - f_s)\right), \tag{4.15}$$

while the common signal gives the perimeter

$$\mathrm{FSR} = \frac{1}{2S}\left((S\,\mathrm{FSR} + f_s) + (S\,\mathrm{FSR} - f_s)\right). \tag{4.16}$$

We can write the measured rotation rate as

$$\Omega = \left(\frac{(S\,\mathrm{FSR} + f_s) - (S\,\mathrm{FSR} - f_s)}{(S\,\mathrm{FSR} + f_s) + (S\,\mathrm{FSR} - f_s)}\right)\left(\frac{\lambda\,c\,S}{4\,A\,n\sin\left(\phi + \theta_T\right)}\right) \tag{4.17}$$

Rotation measured this way has the interesting property that since the beams differ in optical frequency by many MHz, the measurement will not be effected by Adler pulling (see Section 2.3.3). In this way, the alternate-split method provides a method of operation immune from the effects of backscatter. Additionally, if measurements of the regular, single-mode Sagnac frequency are made occasionally then we can obtain a direct measure of the Adler pulling.

4.5.1. Experiment

Because of the excessive time required to achieve a stable single-mode configuration in the UG-3 laser (up to 10 minutes) it is not practical to simply continue restarting the laser until a stable RF or Sagnac beat signal of the desired frequency is achieved. However, we can measure the chaotic spectrum just after startup, calculate the spectral centroid (see Section 3.4) and be sure that the mode configuration of each beam will converge to a single mode at the frequency of the initially measured spectral

centroid. We can then rapidly restart the laser until a desired splitting configuration is achieved.

The experimental hardware and algorithm for determining the spectral centroids has been described in Section 3.5.1. Using this technique it is typically possible to achieve a desired mode configuration in around 5 minutes or less. In this experiment a simple LabVIEW program (running on system named DAQ2 in Figure 4.15) alternated the desired mode configuration between splitting orders of -2, 0 and 2. Splitting orders of -1 and 1 are slightly more probable, however because the Lorentzian peaks observed on the Fabry-Pérot spectra overlap slightly and the noise level is high, they are more prone to misidentification. This is why -2 and 2 were used. The laser was operated for 10 minutes in each configuration.

In order to recover the RF beat signals, a heterodyne technique is used. Separate photomultipliers are used for the $(S\,\mathrm{FSR} + f_s)$ and $(S\,\mathrm{FSR} + f_s)$ respectively. The current output from each is capacitively coupled and mixed directly[1] with a reference signal from individual GPS locked RF oscillators. These mixed signals are then presented to the inputs of two narrow band, AM communications receivers. The pass-bands are set to the frequency of the corresponding RF oscillator. These signals are available as the ordinary audio outputs of the receivers. The oscillators are set to give beat frequencies observed on the receivers of close to the nominal Sagnac frequency in each case. This allows these signals to be phase tracked (see Section 6.6.2) using the same audio reference oscillator (REF1 in Figure 4.15) that is used for Sagnac measurement.

The whole setup is shown in Figure 4.15. Note that separate computer systems are used for control of the splitting order and acquisition. The control signal for the RF transmitter used to excite the plasma is provided by an DAC output on a data acquisition card (in system DAQ2) and is passed through the other data acquisition computer (DAQ1). This allows determination of when the laser is operating at the single-mode power level. This is used together with checks of the amplitudes of the other signals in an implementation of the finite state-machine design pattern (see [95]) to determine the state of the laser and and process the data accordingly.

[1] That this works effectively, given the different impedances and long cable runs can be mostly put down to good luck. Ideally a fast trans-impedance amplifier, followed by a double-balanced mixer should be used but this was not available.

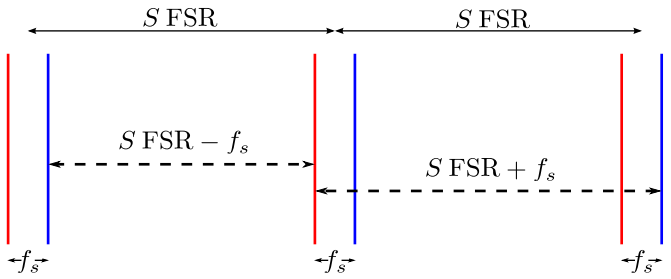


Figure 4.14. Alternate-split mode configurations.

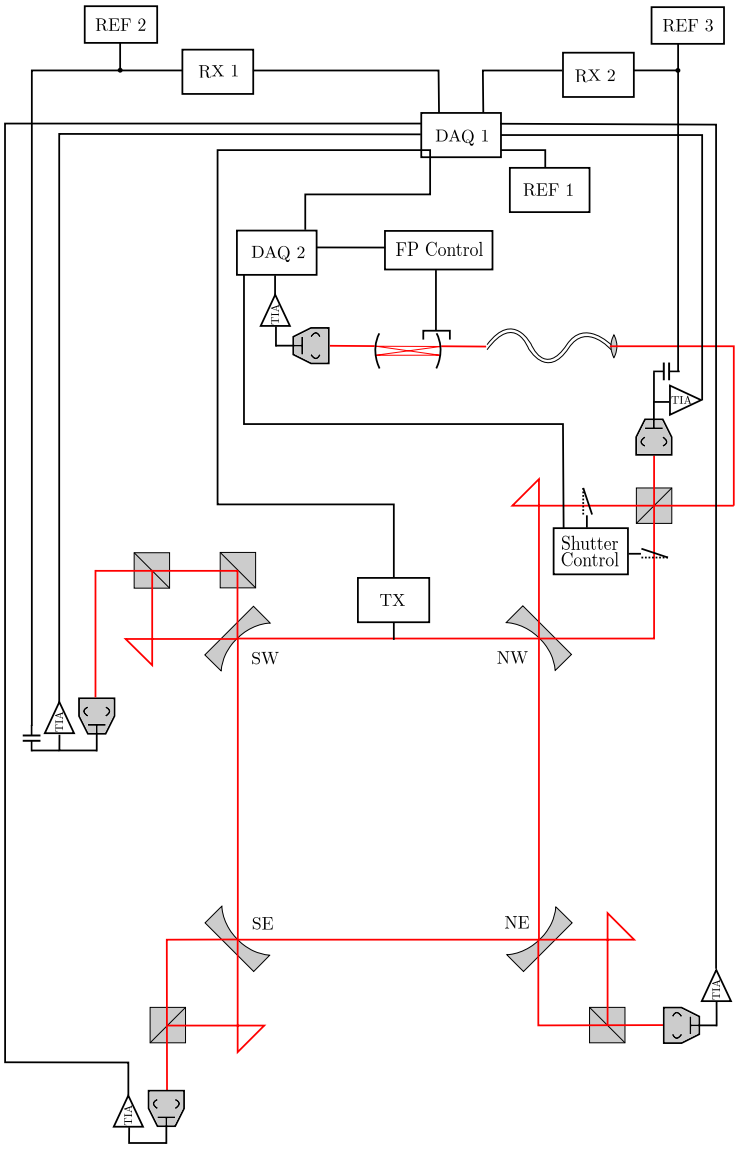


Figure 4.15. Experimental configuration for alternate-split operation.

4.5.2. Results

Figure 4.16 shows the results of this experiment when run for a period of around 3 days during which the cavern was minimally disturbed.

Typically a 10 minute block (comprised of 1 minute averaged measurements) is achieved for each of the three possible configurations per hour. The remaining time is spent restarting the laser to obtain the desired mode, and waiting for mode competition to take effect and stabilise the signal once the desired mode configuration has been determined. Overall this mode of operation gives around a 50% measurement 'duty cycle'.

A 2σ cut-off filter[1] has been applied to the raw measurements. Following this, each block of measurements has been identified and the blocks averaged. A cubic spline was then constructed between these points. A smoothing coefficient of 5×10^{-7} (see [96]) was used and the points were not weighted.

The FSR, and then perimeter can be recovered (Equation 4.16). This is shown in Figure 4.17. The perimeter has been expressed in wavelengths, assuming a wavelength of $632.991\,30 \times 10^{-9}$ nm (weighted natural neon equivalent line centre) and applying a refractive index correction of $1+1.466 \times 10^{-7}$ due to the gas mixture (4.377 mbar total, 4.5 % natural neon). It is also necessary to apply a small correction due to the Guoy phase shift which occurs when a beam passes through a waist. Following Siegman 7.4 [58], the total additional phase per arm is

$$\Phi(z) = \tan^{-1}\left(\frac{\lambda z}{\pi w^2}\right) \tag{4.18}$$

where z is (in the case of UG-3) half the arm length. Because of the astigmatism of the UG-3 beams, the Guoy phase shift is the average of that calculated for in-plane and out-of-plane waists. Table 4.4 shows the w values (in mm) for UG-3. Overall the Guoy phase shift accounts for a reduction in the measured length by 0.9745λ around the whole cavity.

Note that assuming ideal mirrors[2] the number of wavelengths around the perimeter

[1] This eliminates the small fraction of the data points which are clear outliers and usually occur when the software has misidentified the operating configuration.

[2] Note that even if a non-zero phase shift occurs at the mirrors the overall perimeter will still be quantised into integer wavelengths.

	East/West arm	North/South arm
In-plane	0.972	1.481
Out-of-plane	1.823	1.539

Table 4.4. Astigmatism of the UG-3 beams illustrated by the spot sizes (in mm) at the mirrors.

must in reality be an integer to satisfy the resonance condition. We have rounded the perimeter measurements to integer wavelengths and this is also shown in Figure 4.17.

Note that this method of perimeter measurement is clearly sufficiently sensitive to measure the perimeter at the wavelength level. It is interesting to note that if the exact number of wavelengths can be counted then the uncertainty on this count is zero. The true geometric perimeter is related to this count by the refractive index n and the optical frequency, which will move relative to line centre as the perimeter changes. Each of these can potentially be measured independently to further improve the perimeter measurement.

Note that the perimeter shows a clear periodic behaviour with period around 12 hours. This is caused by tidal strain. Note that while tidal strains causes the perimeter to change by one wavelength about every 2 hours, we would not normally expect the laser to jump modes this often when operating in the single mode regime. This is because of the considerable de-tuning afforded by the width of the helium-neon gain curve and the spectral hole-burning phenomena.

The average perimeter is also expressed in meters and given in Table 4.5. The perimeter calculated in this way compares favourably to results from an earlier survey.

The Sagnac frequency has also been calculated in accordance with Equation 4.15. The average value compares favourably with the single-mode Sagnac frequency, although we note that the noise in the measurement is around 2 orders of magnitude larger than the variations measured in single mode. This obscures the tidally induced variations.

The increase in measurement noise is not surprising. Of the signal that is ultimately measured following the heterodyne procedure, only 1 part in 5×10^3 constitutes the calculated Sagnac frequency. This could be improved by using splitting orders of 1 and -1. If high speed data acquisition hardware could be applied to directly track

Alternate-split Sagnac	1512.619552 ± 0.015 Hz
Single mode Sagnac	1512.620884 ± 0.0001702 Hz
Single mode Sagnac (perimeter corrected)	1512.620886 ± 0.0001711 Hz
FSR	3896960.80406 ± 0.04 Hz
Perimeter	$76.92981034 \pm 8.1 \times 10^{-7}$ m
Perimeter according to survey	$76.93973 \pm 1 \times 10^{-2}$ m

Table 4.5. Summary of average measurements comparing single mode and alternate split operation.

the phase of the RF beat signals, then it is likely that the noise in this measurement could be much reduced.

As the calculated value is immune from the effects of backscatter, the difference between these two estimates gives a measure of the frequency pulling/pushing caused by backscatter coupling. The results indicate an anti-phase backscatter modulation pushing the two beams apart by *at most* 1.35 mHz, or about 1 part in 10^6. Already this is relatively small; of the order of the daily variations caused by earth strain. The true value of the pulling is likely to be much less.

Figure 4.17 shows the single mode Sagnac frequency corrected after taking into account the calculated perimeter. Applying this correction we observe a small change in the measured Sagnac frequency of at most 3×10^{-2} mHz, and correspondingly a 0.55 % reduction in the variations as found from the standard deviation. This relatively slight improvement is not unexpected, since the tidally induced tilts dominate the similarly-induced perimeter changes.

Because of this relatively small improvement, operation using the alternate-split method or any other other perimeter correction technique is unlikely to be a useful technique for long-term operation unless the laser is built with the plane perpendicular to the rotation axis.

Note that the error in measurement of Sagnac frequency by the alternate-split method decreases as the scale factor of the laser increases. This is unfortunate given that it is the smaller devices which are more susceptible to backscatter pulling and thus have the most to gain by eliminating it through this technique.

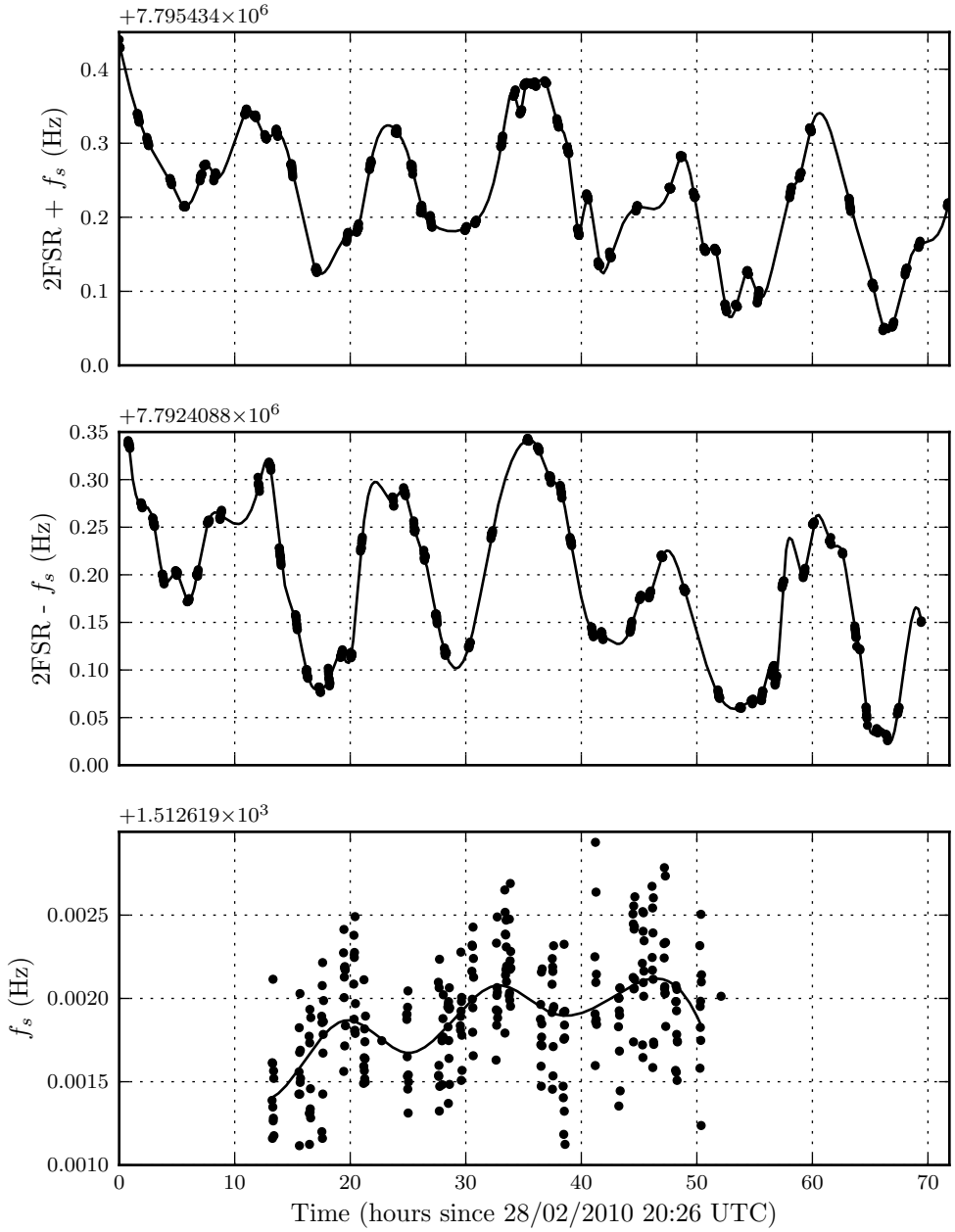


Figure 4.16. Raw results of the alternate-split experiment, with cubic spline interpolation between blocks (solid line). The lower plot (f_s) is the regular Sagnac ($S = 0$), shown here so that the tidal variations can easily be compared.

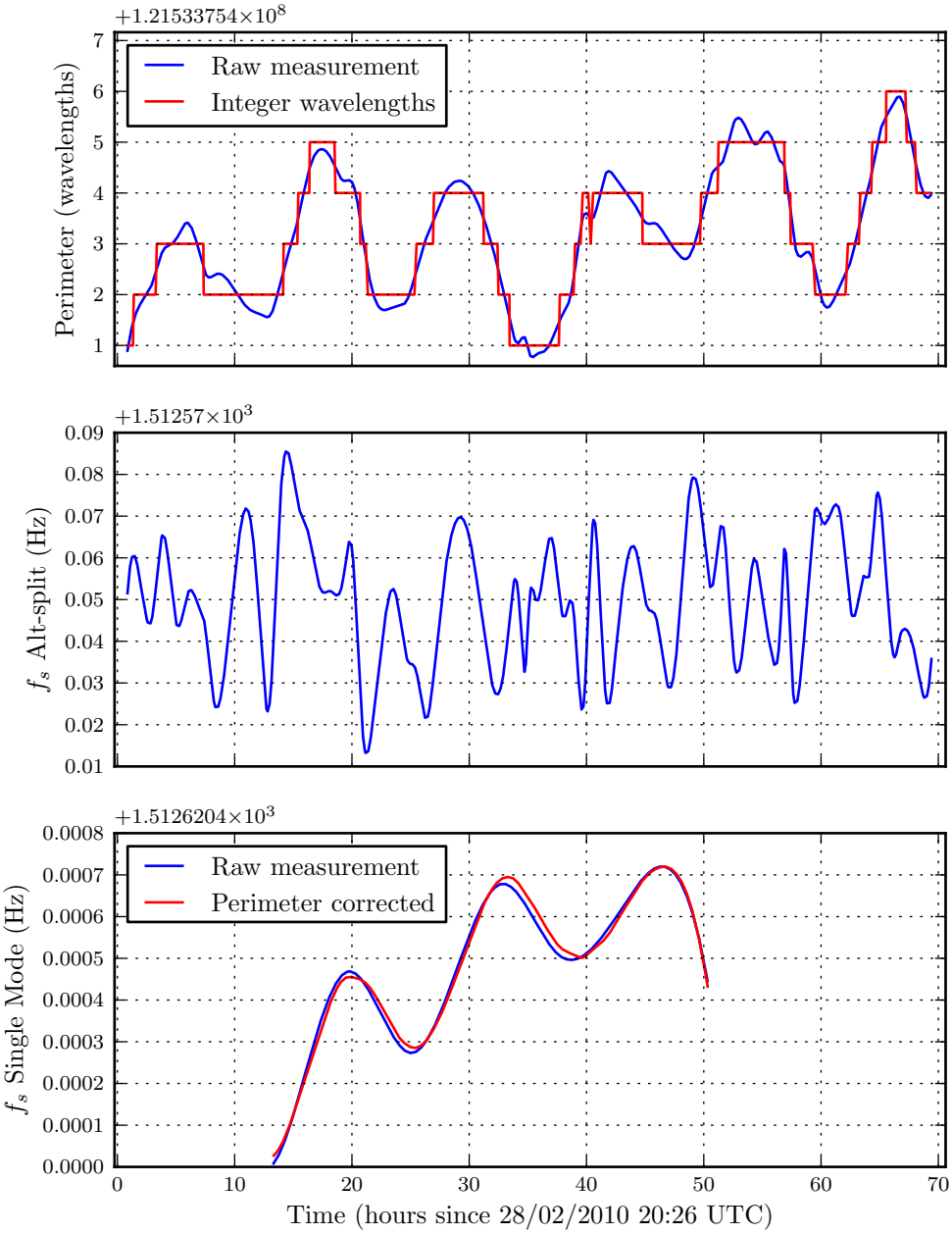


Figure 4.17. Calculated perimeter and Sagnac frequencies obtained from alternate-split operation. Note that the Sagnac frequency determined in this way is not sufficiently precise to show the tidal variations.

4.6. Injection seeding

A significant limitation to alternate-split operation is the time required to achieve a stable mode configuration. This limits the time between measurements to several minutes. However micro-seismic activity perturbs the geometry of the laser on a much faster time scale. Ideally we would like to obtain measurements at a rate of about 5 Hz to fully correct for the changes in geometry. This implies measuring the alternate states for around 100 ms before switching.

One way fast switching might be achieved is through the use of *injection seeding*.

During the normal start-up of a laser, the cavity is initially flooded with a wide spectrum of spontaneously emitted photons. Those with frequency matching a cavity resonance and falling within the gain curve will be amplified. It is only following the slow process of mode-competition that single mode operation can be achieved.

The idea behind injection seeding is to use a secondary laser to artificially provide an initial population of 'seed' photons with the frequency of the desired mode, and vastly outnumbering the population of photons provided by spontaneous emission. Only after the seed laser has provided this population of photons is the main ring laser started. The seed laser is stopped simultaneously. The ring laser should start up in the desired configuration since the population of photons competing for gain has been vastly skewed by the seed population.

4.6.1. Seed beam intensity requirement

In order to investigate the viability of this technique, an experiment was conducted to determine the intensity that would be required of the seed beam in order to successfully seed the operation of a particular mode. This was done by interrupting the RF power to the discharge then slewing the control signal back to the operating level. This technique is detailed in Section 3.5.1 and is illustrated in the portion of Figure 4.18 after 16 ms.

Immediately before and after restarting, scans of the mode structure of each beam were obtained by Fabry-Pérot spectrometer using the procedure detailed in Section 3.5.1. In each case the laser was initially running with a steady single mode in each direction.

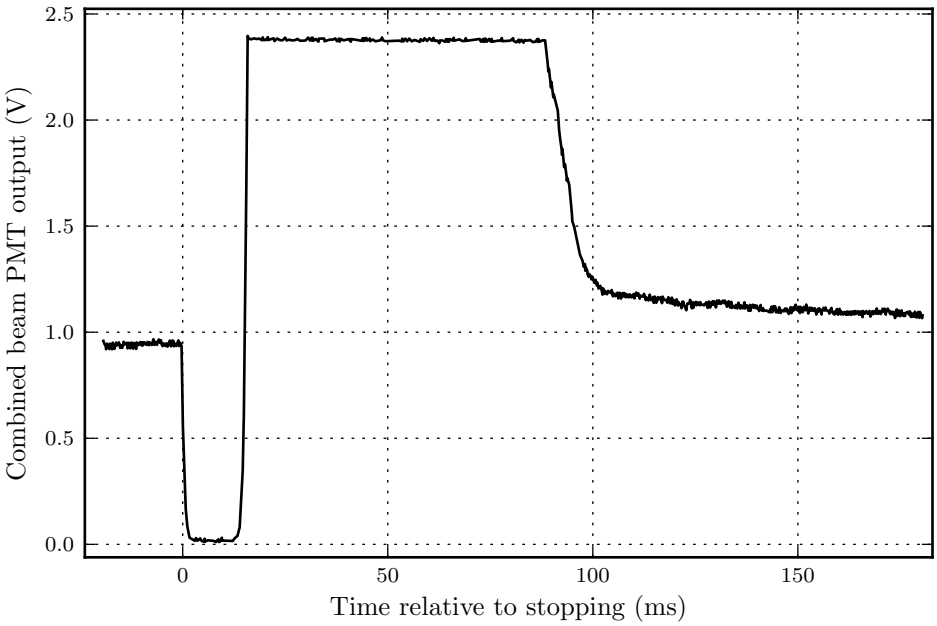


Figure 4.18. Laser output during the interruption-restart procedure.

The time delay t_d between interruption and beginning of the restart procedure was varied from 0 ms to 20 ms. It was found that the original mode configuration and Sagnac signal was restored in all cases when the delay was less than 11 ms (Figures 4.19 and 4.21). When the delay was greater than 12 ms the laser always produced an initially unstable mode configuration (Figure 4.20). After waiting several minutes for the mode pattern to stabilise it was found that the new configuration obeyed the same statistical distribution as when started normally.

The ring down time t_r for this laser is 0.415 ± 0.006 ms. This means that after 11 ms the circulating power has reduced to just 2.9×10^{-12} of the circulating power under steady state operating conditions. The laser power under normal operating conditions was 54.5 nW (single beam), which corresponds to an intra-cavity power P_i of 247 ± 10 mW. This means that after 11 ms the expected number of circulating photons is just

$$\frac{P_i \lambda P}{\exp(t_d/t_r) hc^2} = 6 \pm 3, \tag{4.19}$$

and this number reduces to < 1 when $t_d > 11.75$ ms. This indicates that provided they have exactly the right frequency, only a few photons are needed to successfully

induce seeding. Note that even with the gain switched off, the optical frequency of the circulating beam exactly tracks the resonant condition of the cavity as the length changes. This can be understood in terms of Doppler shifts caused as the mirrors move (see Siegman 25.32 [58]).

We observe that when the power is increased from the normal operating level to above the multimode threshold, it takes of the order of 10's of seconds for the new modes to 'grow' relative to the original mode. Overall this suggests that while initially introduced as a necessity to reliably restart the plasma, the short burst of high power may be beneficial to the seeding process.

Overall, these results indicate that limitations to injected beam power are not likely to present significant problems to an injection seeding experiment. It also suggests a reduced frequency stability requirement because the seed beam need only be maintained at resonance a short time.

4.6.2. Possible implementation

Figure 4.22 details a possible implementation of the injection seeding concept.

A frequency stabilised laser with the same isotope mixture as the main ring laser provides the laser source for the seeding. This is followed by an acoustic-optic modulator (AOM)[1] which splits the beam into two separate beams, separated in frequency by FSR$\pm f_s/2$ (or higher multiples of the FSR). A right angle prism[2] is used to separate the two beams.

The splitting could also be achieved by a rotating radial grating or beam splitter followed by two electro-optic modulators (EOM). The latter arrangement may be more expensive than a single AOM, but eliminates the long throw (at least 2 m between AOM and mirror[3]) and change in angle ($\sim$2.9 mrad assuming a UG-3 sized laser[4]).

Following the splitting, additional optics is then used to precisely align and match

[1] Also called a Bragg cell.

[2] Large prisms of this design are readily available, normally being used for splitting images formed by a single video lens.

[3] Assuming a TeO_2 AOM with longitudinal acoustic velocity 4.2 km s^{-1} [97].

[4] Note that the Bragg angle θ_B induced by an AOM depends on the acoustic velocity v in the AOM crystal, $\theta_B = \sin^{-1}\left(\frac{\lambda\,\mathrm{FSR}}{2v}\right)$ in this case.

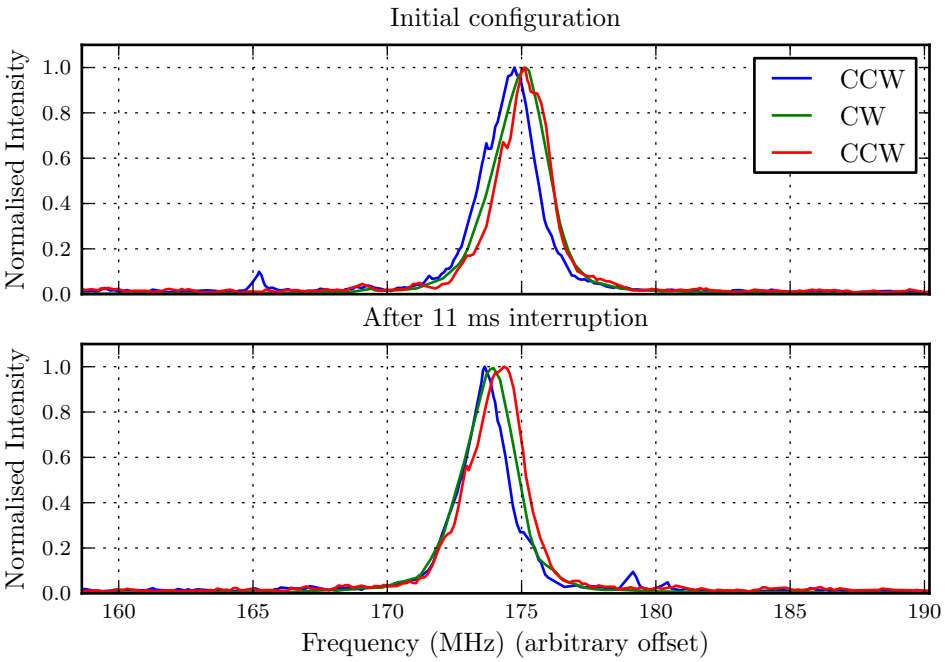


Figure 4.19. Initially single-mode configuration where seeding is successfully demonstrated.

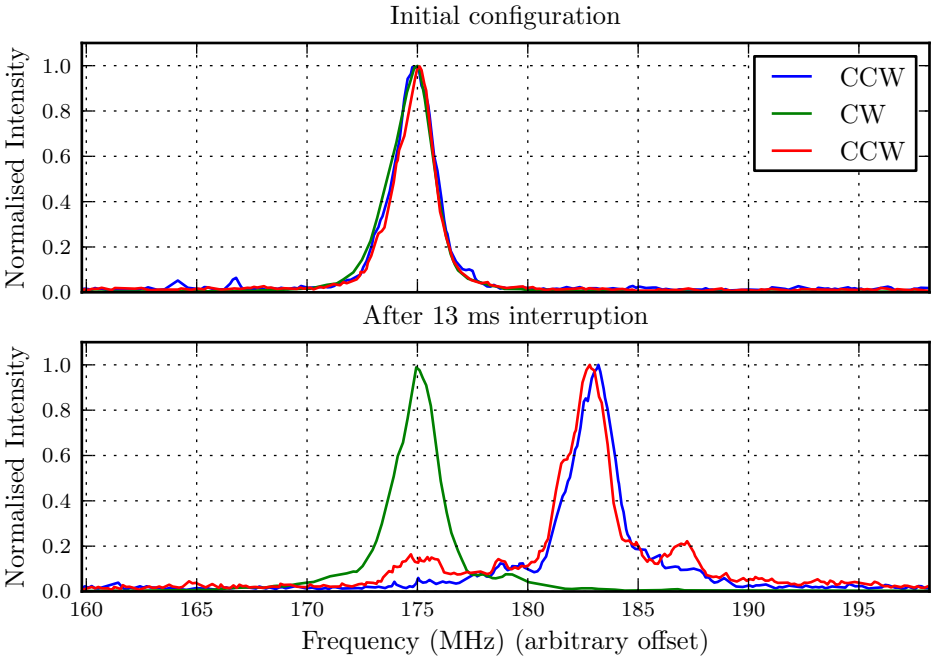


Figure 4.20. Initially single-mode configuration where seeding is not demonstrated.

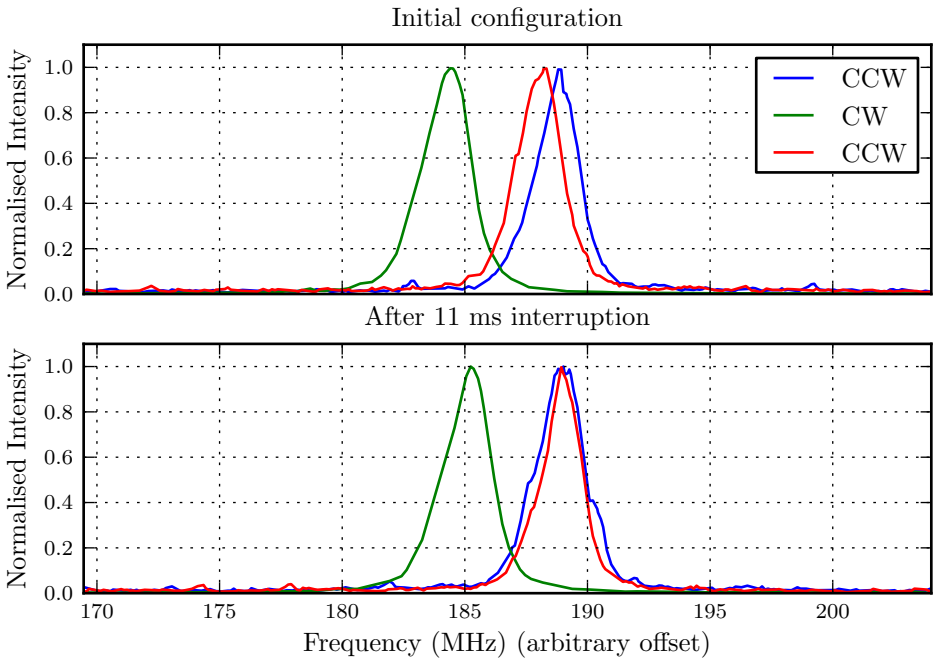


Figure 4.21. Initially split-mode (1×FSR) configuration where seeding is successfully demonstrated.

divergence, spot size and astigmatism of these beams to the combined beams from each of two corners of the ring laser. Fast shutters are provided on one of the beams following the each beam splitter. When in the seeding phase, these shutters will be closed.

At the beginning of the seeding phase the drive frequency of the AOM and offset of the seed laser could be roughly set to give the desired mode configuration inside the ring laser. For example, if FSR+f_s is desired (see right side of Figure 4.14) then the seed laser offset would be set to approximately $+\frac{1}{2}$FSR (relative to line centre) and the AOM set to FSR+f_s.

Sum and difference signals from the photo-detectors (used for Sagnac detection when the shutters are open) provide a convenient means of establishing feedback to tune the seed beams to the ring cavity resonance. During injection, the sum signal can be maximised through feedback to the seed laser offset. When maximised, the average beam frequency will be half way between resonances of the ring cavity. The difference signal can then be minimised by controlling the drive frequency of the AOM. The difference signal is minimised when the splitting is an integer multiple of the ring cavity FSR.

With both feedback loops optimised, the seed beams are exactly on a resonance of the ring cavity. The seed laser can then be blocked then the ring laser restarted in the desired configuration.

Achieving successful seeding with an external beam requires the ability to control the optical frequency of the seed beam to high precision. The overall optical frequency stability requirement can be determined from the width of the ring cavity resonance,

$$\text{FWHM} = \frac{1}{4\pi t_r} \tag{4.20}$$

which is 192 Hz for the UG-3 laser. In practice we would expect good results if we could maintain optical frequency at the ∼100 Hz level for at least the length of any switching delays, say around 10 ms.

Extrapolating from Allan variance statistics given for a popular model offset-locked iodine-stabilised laser (Winters model 200) we find that a stability of around 400 Hz should be possible over these time scales. Frequency stability could probably be enhanced during the injection if an external signal from the ring cavity could be

provided to the feedback system of the offset laser.

4.7. In-ring-down Sagnac measurement

Let us consider a ring laser operating using the alternate-split technique at a high rate, made possible through the use of injection seeding. It is interesting to note that such a laser will spend a considerable fraction of time in the ring-down phase.

The Sagnac frequency splitting that occurs during ring-down is different from the regular 'steady-state' Sagnac in a number of interesting ways.

1. It is free from pulling due to dispersion, which inevitably accompanies laser gain.
2. It is free from Sagnac frequency shifts caused by alternate displacement of the gain curves by plasma flow.
3. The circulating beams are not continually perturbed by spontaneous emission, they are only effected by the spontaneous emission present in the cavity before the switch off.

4.7.1. Experiment

An experiment was undertaken to determine how well the Sagnac frequency can be measured during ring down and to measure any difference between this and the regular steady-state Sagnac.

Power to the UG-3 ring laser was initially under the control of an embedded servo system. Power was interrupted by directly shorting the transmitter control to ground via a software controlled mechanical relay. Starting just before interruption, signals from the combined beams from all of the corners and a GPS locked reference (set to the nominal Sagnac frequency f_r) were acquired by a data acquisition system at a rate of 50 kHz per channel. The reference waveform is necessary to establish a source of accurate timing, independent of the acquisition clock.

4.7.2. Analysis: a general least-squares approach

The time values at each measurement point were obtained by constructing a cubic spline from the reference waveform, finding the zero crossings using Brent's local

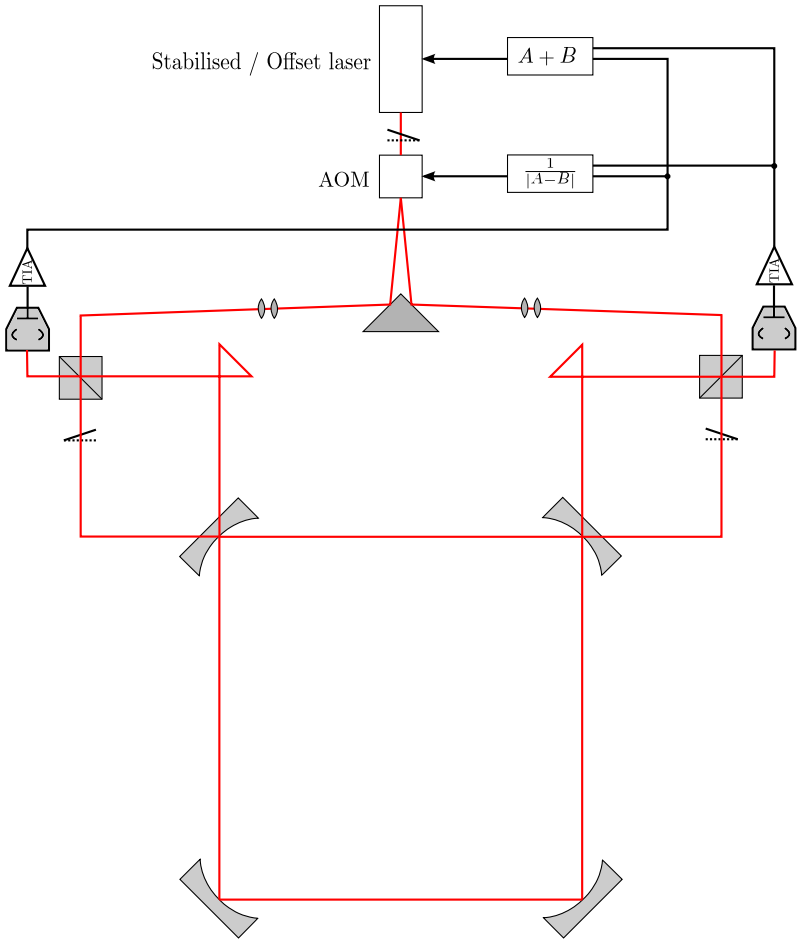


Figure 4.22. A possible optical arrangement for a ring laser operating using the alternate-split technique with injection seeding.

minimisation method [98] and then interpolating linearly through these points to obtain a re-sampled time series.

The least-squares method (minimised using the Levenberg-Marquardt algorithm) was applied to the data and model of the ring-down process,

$$y(t) = \begin{cases} a\sin(\omega_1 t + \phi_1) + d & \text{if } t < t_0 \\ \exp(-\frac{t-t_0}{t_r})\, a\sin(\omega_2 t + \phi_2) + d & \text{if } t \geq t_0 \end{cases} \tag{4.21}$$

where a is the amplitude, ω_1 the steady state Sagnac angular frequency, ω_2 the 'in-ring-down' Sagnac angular frequency, ϕ_1 and ϕ_2 the respective phase offsets, d the 'dark offset' and t_0 the time of interruption.

Parameters d, a and ϕ_1 were constrained during the fitting. The values of these parameters were found from a subset (denoted R) of the data with length of the nominal Sagnac period and beginning at the start of the original time series. These were

$$a = \frac{\max(R) - \min(R)}{2}, \tag{4.22}$$

$$d = \min(R) + a, \tag{4.23}$$

and

$$\phi_1 = \begin{cases} \sin^{-1}\left(\frac{R(0)-d}{a}\right) & \text{if } R(1) > R(0) \\ -\sin^{-1}\left(\frac{R(0)-d}{a} + \pi\right) & \text{if } R(1) \leq R(0). \end{cases} \tag{4.24}$$

Initial guesses of ω_1 and ω_2 were obtained by first minimising just the $t < t_0$ part of Equation 4.21. ϕ_2 was initially set to ϕ_1 and initial guesses for t_r and t_0 were entered manually from experimental values as they should change little between data sets.

An example plot of a ring-down and associated fit to Equation 4.21 is shown in Figure 4.23. This appears to shows a good fit, and in the best case gives a in-ring-down Sagnac estimate within 12 Hz (0.8%) of the steady-state Sagnac.

This is conceptionally the simplest approach, however it was found to not be very robust and often required manual tuning of the initial guesses to get the fit to converge.

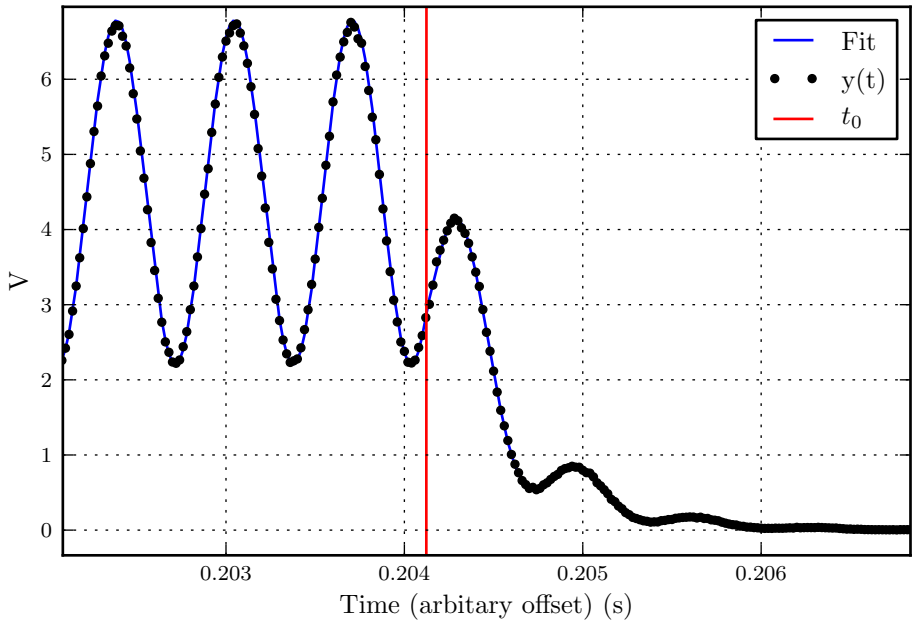


Figure 4.23. A least-squares fit to the in-ring-down Sagnac frequency model.

.

4.7.3. Analysis: an alternate approach

An alternative, more robust algorithm for determining the in-ring-down Sagnac frequency was developed. The basic idea is to transform the time series into a form which does not show the exponential decay, then construct a cubic spline from transformed data. The cubic spline can then be sampled to determine the exact phase corresponding to each peaks and trough. The same approach can be applied to the reference waveform (known frequency) and the in-ring-down frequency derived from the relationship between the phases of the peaks and troughs of the signal and reference. The computational details can be summarised as follows:

1. Obtain $y(t)$ from the raw data, selecting a time series beginning 500 points before the signal drops below 0.5% of the steady-state mean level. Note the t here is the integer sample number.

2. Let

$$m = \text{smooth}(\log(y)) \tag{4.25}$$

where smooth() performs a convolution of a scaled Hanning window function (length 3 × nominal Sagnac period) with the input. The signal is prepared by introducing reflected copies of itself (with the window size) to both ends so that transient parts at the beginning and end of the original signal are minimised.

3. Given:
$$s(t) = \begin{cases} y_0 & \text{if } t < t_0 \\ y_0 - (t - t_0)g & \text{if } t \geq t_0 \end{cases} \tag{4.26}$$
Find t_0, g and y_0 by Levenberg-Marquardt such that $(s(t) - y(t))^2 \sim 0$.

4. Let:
$$n(t) = \exp(\log(y(t)) + y_0 - s(t)) \tag{4.27}$$
and
$$n'(t) = \text{smooth}(n(t)) \tag{4.28}$$
following the definition of point 2, but with a window length of $\frac{1}{2}$ × nominal Sagnac period.

5. Find an array of points $P[n]$ such that t satisfies
$$0 \sim \frac{d}{dt}(\text{spline}(n'(t))) \tag{4.29}$$
and similarly an array $P_r[n]$ based on the reference generator waveform R.
$$0 \sim \frac{d}{dt}(\text{spline}(R(t))) \tag{4.30}$$
where spline constructs a cubic spline from the given data.

6. The frequency is then found from the gradient of a linear fit through the points in question,
$$\frac{\text{gradient}(P, n)}{\text{gradient}(R, n)} f_r. \tag{4.31}$$

After analysing many measurements, this approach was found to be robust enough to be operated in an automated fashion and the overall accuracy of the frequencies before and after cutoff was found to be 1.6% for a single ring-down.

While this does demonstrate the concept of automated in-ring-down measurement, the results are not sufficiently accurate to measure any systematic difference in the

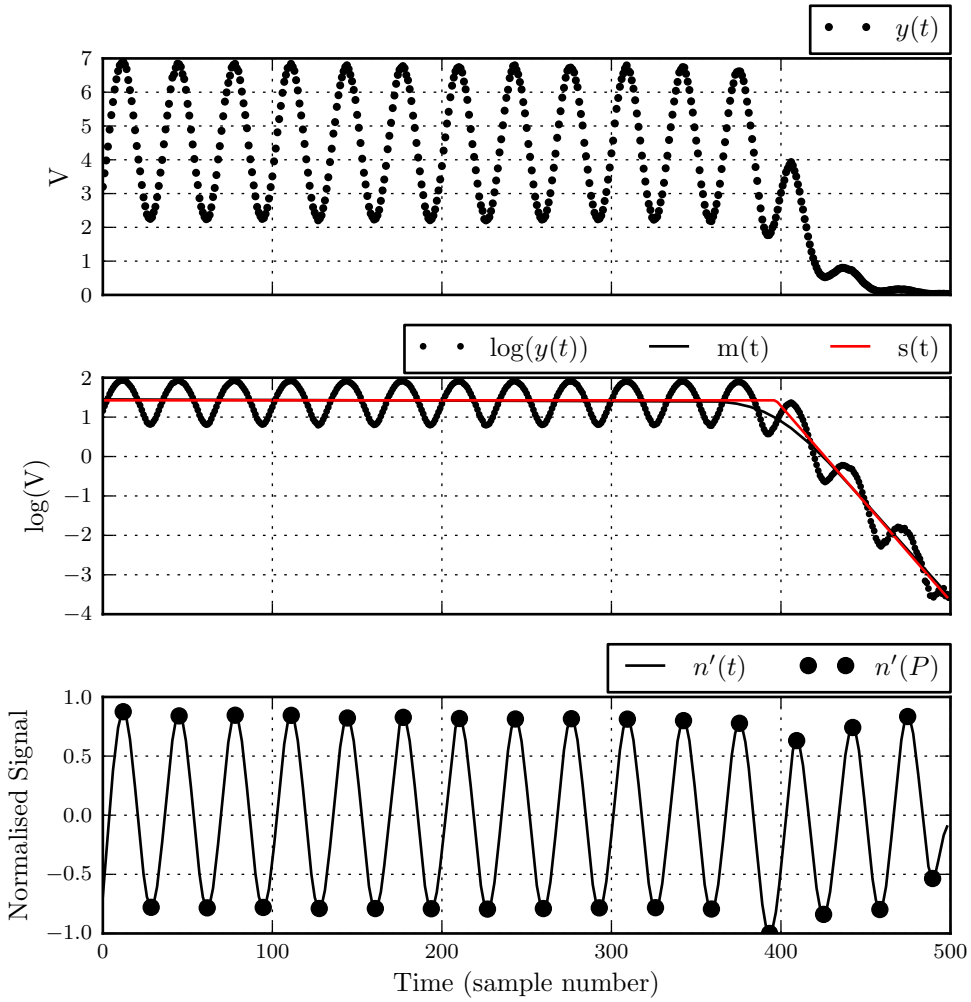


Figure 4.24. Illustration various functions generated processing a ring-down with the method frequency determination method detailed in Section 4.7.3.

Sagnac frequency before and after cutoff. However if applied to an instrument operated using fast injection seeding (Section 4.6), thousands of measurements could be made per hour and the long term average in-ring-down Sagnac frequency could well be competitive with traditional methods.

5. Ring Laser Applications

The purpose of this chapter is to serve as a repository for a number of otherwise unrelated topics that were investigated during the completion of this thesis. In general they all represent possible applications of ring laser gyros. This includes preliminary results from experimental data (microseismic activity, earthquakes), overviews of the practicality of potential measurements (tidal rotations and precession effects) and applications of ring laser gyros in gravity wave detectors.

5.1. Observations of microseismic effects

The microseismic background is a continuous semi-periodic signal comprised of small seismic oscillations and is typically associated with ocean wave activity. Microseisms are observable directly on ring laser rotation signals and also as linear displacement/acceleration signals on sensitive seismometers. Their amplitude and frequency varies daily quite considerably, although they usually fall with the frequency range of 50 mHz to 300 mHz. The precise frequency corresponds to ocean waves of half the microseismic frequency.

The most cited description of the microseismics was presented in the landmark 1950 paper by Longuet-Higgins [31]. In this work they found that:

1. Microseisms originate from surface pressure oscillations caused by the interaction between oppositely travelling components with the same frequency in the ocean wave spectrum.

2. These pressure oscillations generate seismic *Stoneley waves* on the ocean floor. These are boundary waves, similar to Rayleigh waves, but they originate directly at (and propagate along) the boundary between a fluid and elastic media (the Rayleigh wave by comparison originates through interference of other waves at the interface). In both cases, the velocity of the wave does not depend

on frequency [99].

The final amplitude of microseismic waves depends on the ocean wave frequency and the water depth. The theory presented by Longuet-Higgins [31] can be used to estimate the amplitude of vertical displacements due to microseisms given ocean wave data, and this has been tested by Kedar *et al.* [32]. Modelling of the displacements in this way is now sufficiently well developed to allow the origin of microseismic waves to be mapped given data from a network of seismometers. This has interesting applications, for example Kedar *et al.* [32] were able to confirm the reduction in sea ice coverage in the north Atlantic from an associated reduction in microseismic amplitude.

At the present time it appears to be the vertical displacements caused by microseismic waves that have been of primary interest to the geophysical community. Indeed, it is only these displacements and tilts which would be expected to be directly observable from a simple model of the Stoneley/Rayleigh[1] wave propagation.

Microseismic waves are easily observable on all our ring laser gyros. From the data gathered on the G-0 and UG-3 ring lasers, autocovariance periodograms have been routinely computed with the precise frequency of the dominant microseismic motions computed according to the procedure outlined in Section 2.4.3. Interestingly, we see a clear correlation in the period of the dominant microseismic motions as recorded on these two instruments (Figure 5.1). This is interesting because these instruments are located in the same laboratory yet measure rotation about axes at right angles to each another.

One possibility for this correlation is both instruments are simply responding to the same tilt signals associated with the Stoneley/Rayleigh wave. A further study, making use of a precision tilt-meter[2] and the vertical seismometer observations should be able to determine the extent to which this is true.

An alternative, and perhaps more interesting explanation is that they are both observing true rotations. These could be excited via a cross-coupling of the Stone-

[1] Based on the description presented by Kaufman [99], the term Stoneley wave strictly applies only when the wave is travelling on the ocean floor.

[2] While such an instrument is available in the Canterbury ring laser group, it was not operational at the time the data presented in Figure 5.1 was gathered.

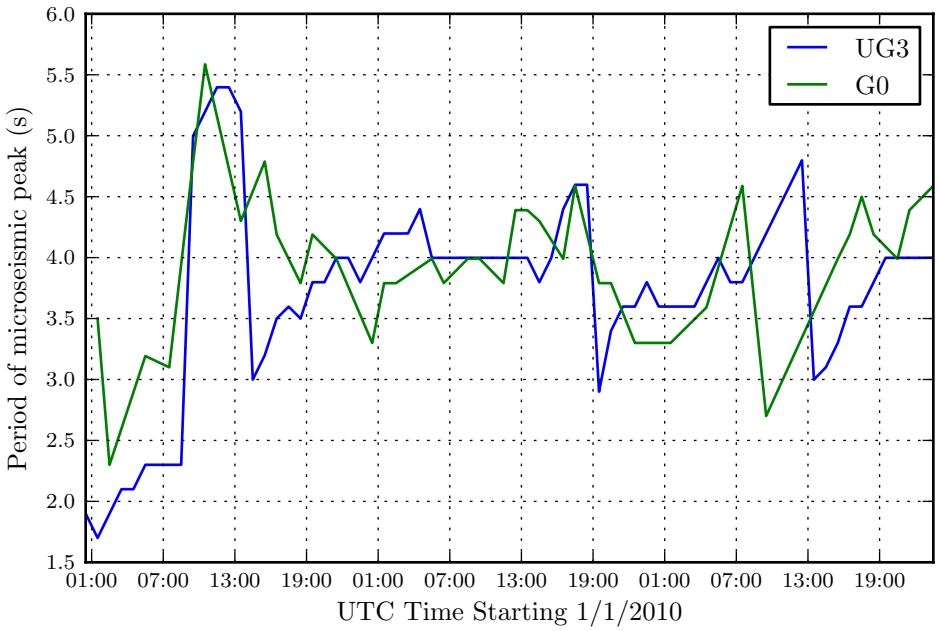


Figure 5.1. Period of dominant microseismic motion recorded in two orthogonal planes.

ley/Rayleigh waves into the two orthogonal degrees of freedom. Such coupling mechanisms can arise via lateral heterogeneities and are a well studied problem in geophysics. For example Maupin [100] has outlined how a Rayleigh wave, incident on a boundary can be partially converted into Rayleigh overtones and Love modes, the latter of which would be expected to produce a rotation signal on a ring laser gyro parallel to the surface.

Should such a mechanism prove to be responsible, and should it prove possible to obtain the cross coupling coefficient for the two degrees of freedom, then it is reasonable to hypothesise that a characteristic distance to the source might be obtained from the ratio of the intensities of the coupled modes in each degree of freedom.

5.2. Earthquake induced permanent surface deformations

The modern understanding of earthquake motions relies on the concept of a *rupture zone* beneath the earth which undergoes a sudden displacement. The *scalar seismic moment* M_0 is a measure of the total amount of energy transformed during an

earthquake. It is defined as

$$M_0 = \mu A D \tag{5.1}$$

where μ is the shear modulus of the rocks involved in the earthquake (typically 32 GPa or 320 G dyn cm^{-2} in the crust [101]), A is the area of the rupture zone and D is the average displacement of the zone. The *moment magnitude* M_W which is typically reported for earthquakes can be found directly from the scalar seismic moment using the expression [102]

$$M_W = \frac{2}{3} \log_{10} M_0 - 10.7 \tag{5.2}$$

which has units of energy (in ergs, note that by convention CGS units are usually used in this type of work).

By way of example, a typical earthquake of around magnitude 3 corresponds to an area of about 100 m^2 moving 0.1 m (or smaller area and correspondingly larger displacement). It is clear from this illustration that earthquakes can cause permanent deformations of the surface near to the rupture zone in addition to the travelling seismic waves we more commonly associate with them.

Okada [103] has reviewed and presented analytical solutions for the deformations that would be expected for different earthquake geometries. Direct measurement of a permanent rotational deformation would represent an important clarification of the underlying theory.

5.2.1. Co-seismic rotations

In this analysis we consider the special case of small[1], shallow[2] ruptures where the surface of the rupture zone is orthogonal to the surface of the earth and the displacement is parallel to the surface of the earth. These are known as *strike-slip* ruptures with *vertical dip* and represent an ideal case for measuring the rotation with a ring laser parallel to the surface of the earth.

Analytical solutions for the displacement field from an event of this type were presented by Okada (Equation 8) [103]. After making some simplifying assumptions, neglecting the exact geometry of the rupture zone and observer, the displacement of

[1] Dimensions of the rupture zone are small compared to the observation distance.

[2] The depth of the rupture zone is small compared with the horizontal distance to the observer.

the ground surface parallel to the fault strike can be shown to be

$$U_x = -\frac{U_1 A m}{2\pi y^2} \tag{5.3}$$

where U_1 is the rupture zone displacement, A is the rupture zone area, y is the horizontal distance from the fault to the observation point, U_x is the movement of the observation point parallel to the fault strike and m is defined as

$$m = \frac{\mu}{\Lambda + \mu} \tag{5.4}$$

with Λ and μ the Lamé constants of elasticity for the medium (sometimes given using the alternate symbols l and m). These are derived from the Young's Modulus and Poisson's Ratio. Given a Poisson's ratio of ~ 0.3 (see [104]), m can be shown to be ~ 0.4.

It is possible to find the rotation from half the curl of the displacement field. Substituting the numerical values allows us to give an expression for the rotation angle θ in nrad,

$$\begin{aligned} \theta &= 0.064 \frac{U_1 A}{y^3} \\ &= \frac{0.027}{y^3} \exp(3.45\, M_W) \end{aligned} \tag{5.5}$$

with y in km.

Figure 5.2 shows a plot of the expected rotation angles as a function of distance to the epicentre and quake magnitude. Note that the rotation angle has an inverse-cube dependence on the distance. For this reason only relatively nearby earthquakes produce potentially measurable rotations.

Figure 5.3 shows the frequency of potentially observable events given the rotation angle resolution. This was calculated using historical earthquake data[1] from 1/1/1990 to 1/6/2010 and Equation 5.5. Distances to the Cashmere cavern were calculated using the Haversine formula.

5.2.2. Analysis of a nearby magnitude 3.8 earthquake

This section concerns the analysis of a nearby earthquake (GeoNet identifier 3285425, location 43.24669°S, 172.7632°E) that was recorded on 3/4/2010 at 19:10:27 UTC.

[1] We acknowledge the New Zealand GeoNet project and its sponsors EQC, GNS Science and LINZ, for providing data used in this study.

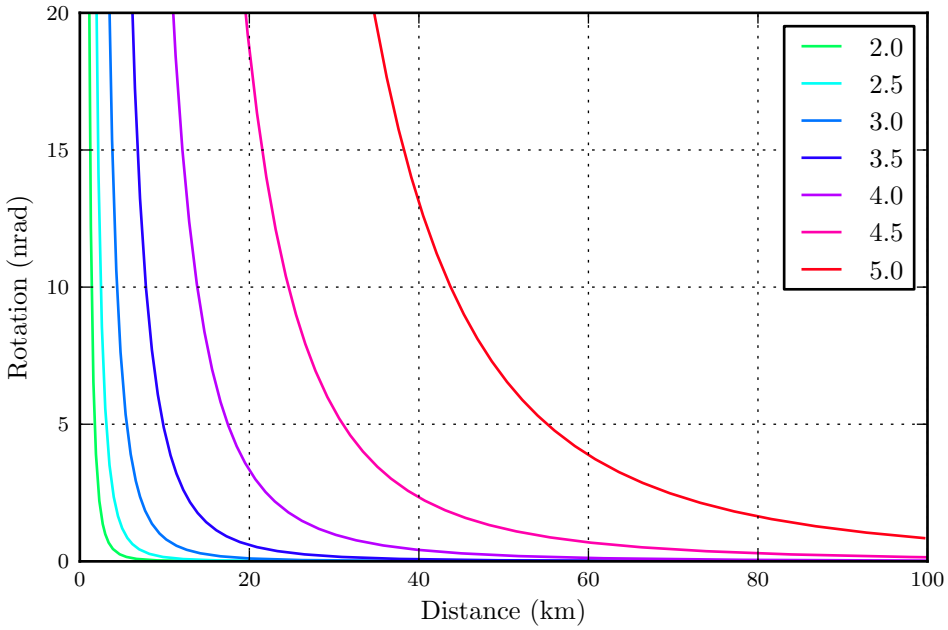


Figure 5.2. Co-seismic rotation due to earthquakes of varying magnitude (indicated by the curve colour).

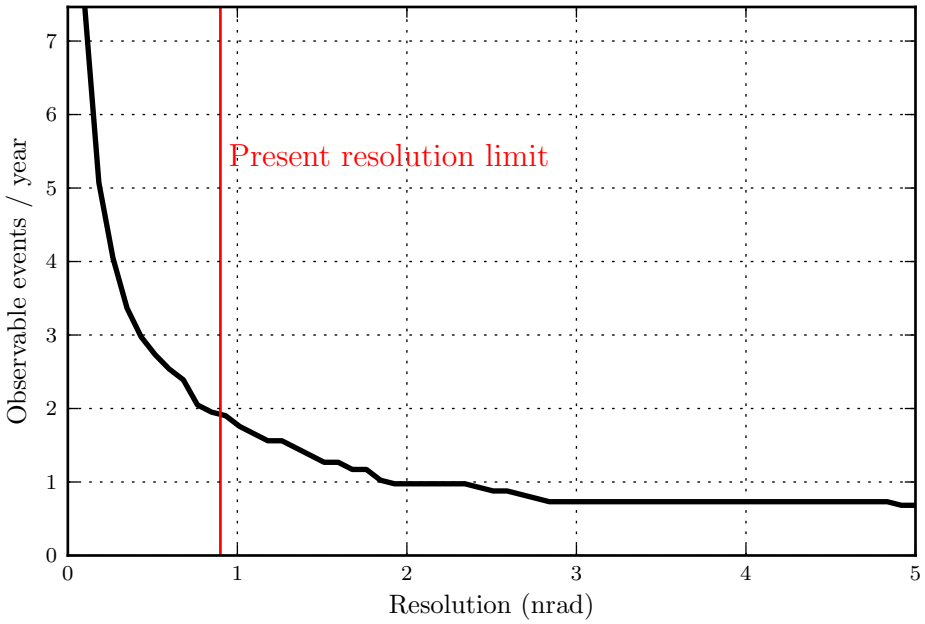


Figure 5.3. Observability of co-seismic rotations based on historical seismic data.

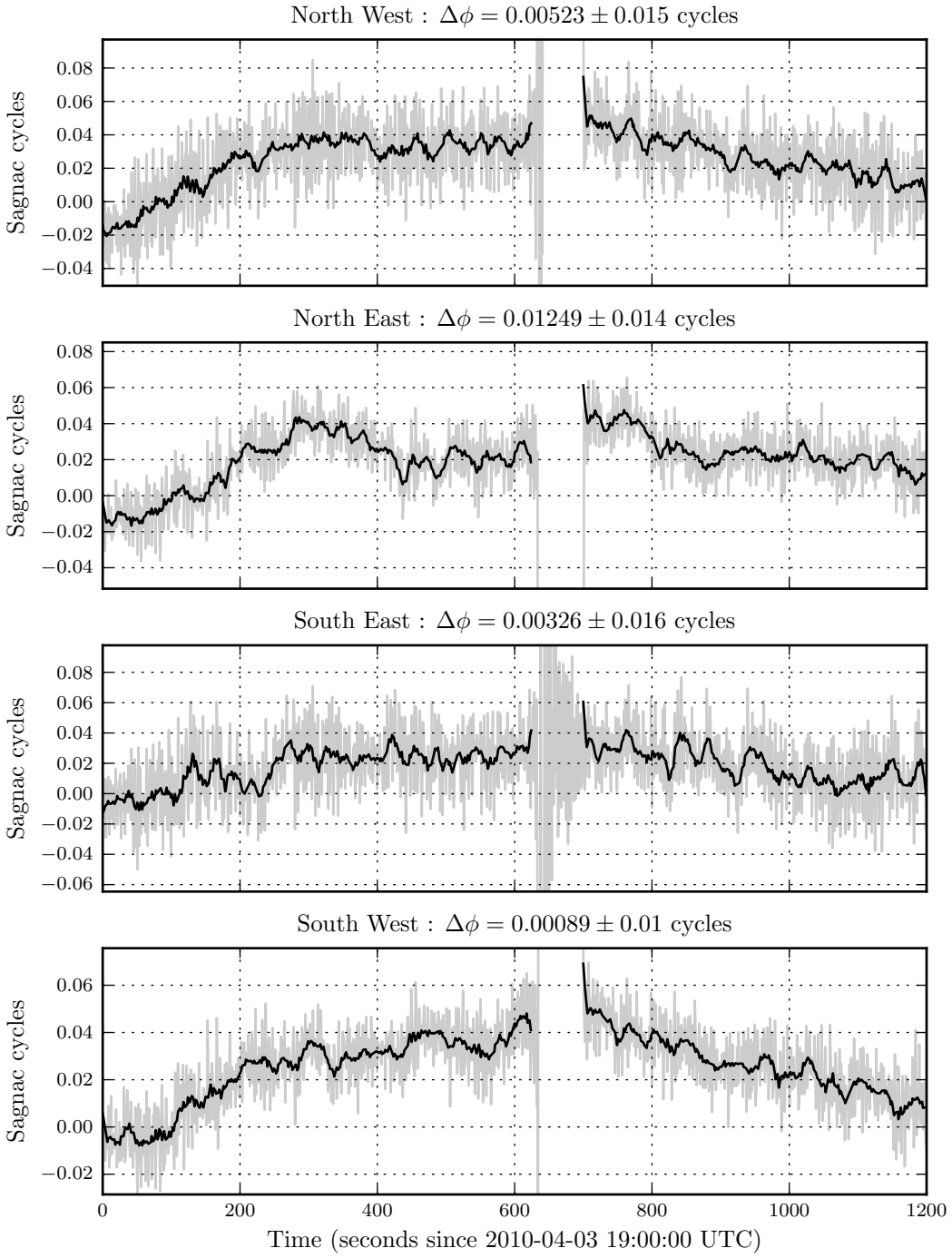


Figure 5.4. Sagnac phase shifts measured on UG-3 for a nearby earthquake. Dark line is the raw data (grey) smoothed with a 10 s moving average. Phase differences ($\Delta\phi$) and uncertainties (1σ) reported are from 200 s portions before and after the quake. The average phase shift (0.0055 ± 0.0070 cycles) is not statistically significant.

This quake had a recorded magnitude of 3.825±0.2 and the epicentre was 38 km NNE (17.3°) of the Cashmere cavern at a depth of 33 km.

Raw accumulated phase data (see Section 6.6.2) from each corner of UG-3 was first de-trended by fitting a straight line to the portions of data before and after the event and subtracting the average fitted line. In most earthquakes the phase tracking will skip an integer number of cycles during the earthquake. This was corrected and the resulting data from each corner of UG-3 is shown in Figure 5.4.

A permanent rotation should be observable as a shift in the accumulated phase following the earthquake. Portions of data of length 200 s were averaged before and after the earthquake. No statistically significant phase step was observed for this earthquake. Given the standard deviations of the portions of data, the limiting resolution for measuring the phase step is around 0.02 cycles. This means a rotation of greater than 0.9 nrad should be observable.

Substituting the numbers for this earthquake into Equation 5.5, the expected rotation is at most 0.8 nrad which is unfortunately just below the noise floor so we were unlikely to be able to observe any permanent rotation for this event.

As we can see from Figure 5.3, given the noise level in this data set we would expect observable events to occur perhaps once or twice per year. Bearing in mind that we have made a number of simplifying assumptions and not all quakes would be of the right type for causing a permanent rotation, it seems that in order to make a successful measurement we require around 5 years of continuous observations.

Because the primary noise source for short periods before and after earthquakes is the micro-seismic background, their appears to be little hope for improving the measurement resolution. However, given the semi-periodic nature of the microseismic background it may be possible to develop an adaptive filter technique to improve the measurement resolution for this type of measurement.

5.3. Rotations from solid earth tides

Solid earth tides are the deformations of the earth caused by the gravitational pull from the sun and moon, analogous to the more familiar ocean tides. These are very well studied in geophysics [92, 105, 106] and have been measured in New Zealand

since the 1970's [107]. The signals are quasi-periodic and have both diurnal and semidiurnal components. The differential tidal forces cause well known strain and tilts signals which are regularly observed on ring laser gyros, and are a significant source of noise.

In this section we consider the rotational component which has not (to the authors knowledge) been considered before. In particular, we consider the possibility of measuring this component with a ring laser gyro.

The surface displacements at a point due to the lunar and solar gravitational fields can be computed following Section 7.1.2 of the IERS Conventions (2003) [108]. This is very involved, however fortunately various implementations already exist. In this work we have made use of the Fortran program *solid* by Dennis Milbert.

Strain in a particular direction can be found by first finding the displacements from the reference ellipsoid due to earth tides of two points a_{xyz} and b_{xyz}. These define a line of length d_{ref} which passes through the point of observation (referred to in the program as the observatory). The strain S is then

$$S = \frac{\sqrt{(a_x - b_x)^2 + (a_y - b_y)^2 + (a_z - b_z)^2}}{d_{\text{ref}}} - 1 \tag{5.6}$$

The rotation can be found from the curl of the displacement field, or given the same points as above, the change in angle Ω can be found

$$\Omega = \tan^{-1}\left(\frac{b_y - a_y}{b_x - a_x}\right) \tag{5.7}$$

which can then be differentiated with respect to time to find the rotation rate.

Tilt can be found by first finding the displacements from the reference ellipsoid for three points a, b and c centred at the observer, equally spaced in angle to from an equilateral triangle. These points define a plane from which the projection of the normal vector

$$\mathbf{N} = (a - b) \times (c - b) \tag{5.8}$$

and the vector

$$\mathbf{B} = \begin{bmatrix} \sin\theta \\ \cos\theta \\ 0 \end{bmatrix} \tag{5.9}$$

can be found to give the angle of tilt

$$\theta_T = \cos^{-1}\left(\frac{\mathbf{N}\cdot\mathbf{B}}{|\,\mathbf{N}\,||\,\mathbf{B}\,|}\right). \tag{5.10}$$

The choice of the reference length d_{ref} is largely arbitrary, it is simply necessary to make this large enough to avoid any numerical rounding errors. Typically 100 km has been used.

The above equations have been implemented in the *pysolid* program, usage details for which are available in Section A.3. This module was written to wrap the Fortran program *solid*, and provide a convenient programming interface and ability to calculate strains, tilts, rotations and rotation rates. Because accuracy is important, this program makes use of Vincenty's 'direct' formulae [109] for making the coordinate transformations given a specified reference ellipsoid[1].

Figure 5.5 shows an example plot of the rotational and tilt deformation due to solid earth tides. Figure 5.6 shows the theoretical deviation to the Sagnac frequency of UG-3 due to these effects. Here we see that the Sagnac frequency deviation due to rotation is approximately an order of magnitude smaller than the corresponding deviation due to tilt. While these two signals have a different form, it can be seen that the combined signal is almost impossible to distinguish from the direct tilt signal. Because of this, isolating the geophysical rotation signal is likely to be a very difficult task for an instrument sensitive to tilts like UG-3. However, an instrument built with the plane perpendicular to the earth rotation axis may prove more promising.

5.4. Lense-Thirring precession

Lense-Thirring precession, also known as *frame-dragging* is a predicted consequence of general relativity. It predicts that a gyroscope in the vicinity of a rotating massive body will acquire a small precession rate proportional to the rotation rate of the body and its moment of inertia. The direction of rotation is the same as that of the massive body.

It is possible to linearise the equations of general relativity for small perturbations and weak fields, giving 3-vectors which obey equations analogous to Maxwell's equa-

[1] Reference ellipsoids WGS 84, GRS 80 and WGRS 80/84 are all available however in this work the default, WGS 84 is assumed.

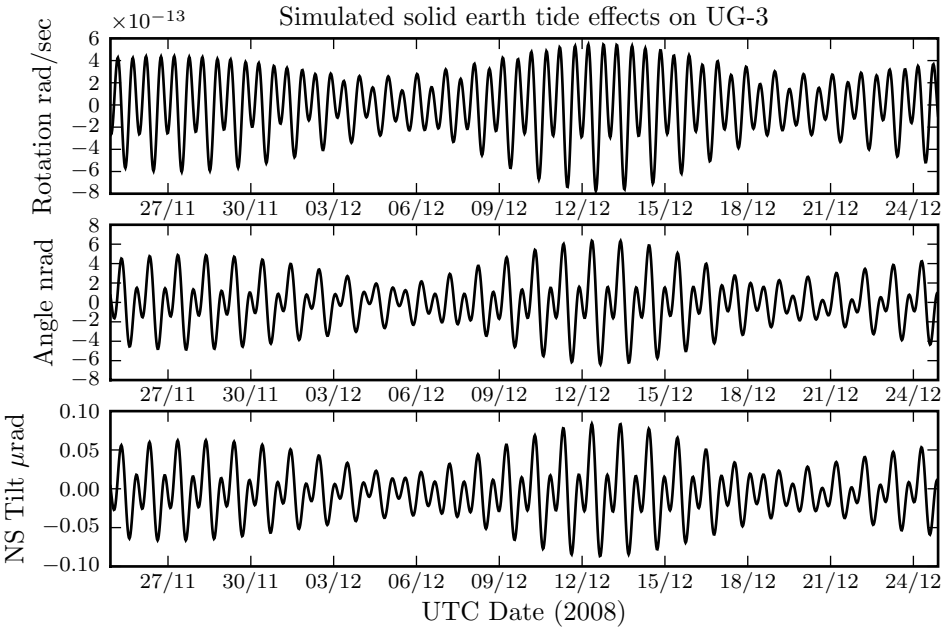


Figure 5.5. A time series showing rotational and tilt deformations due to solid earth tides.

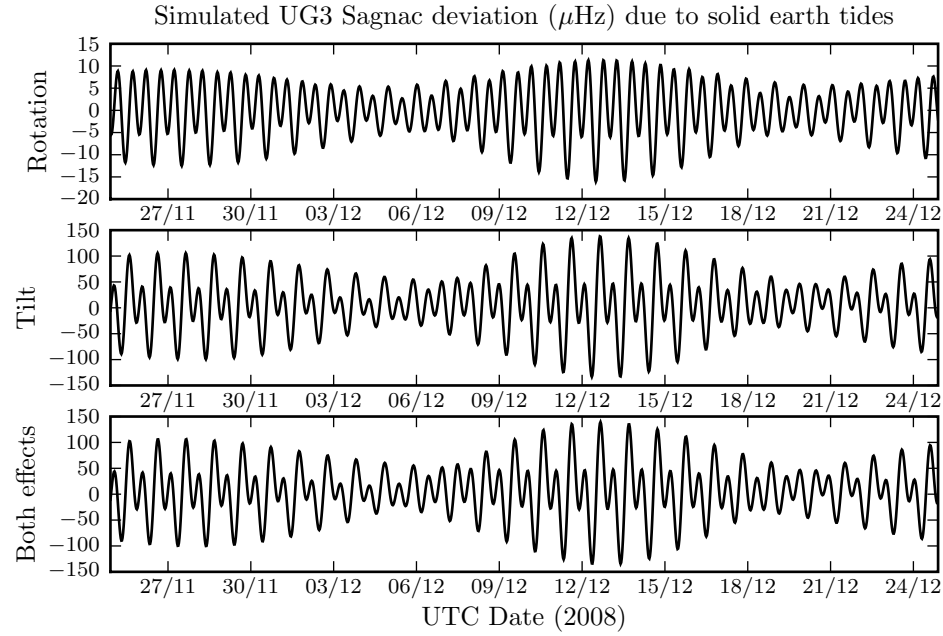


Figure 5.6. Sagnac deviation effect on UG3 of the solid earth tide effects shown in Figure 5.5.

tions for the electric and magnetic fields. The GEM equations [110, 111] of *gravitomagnetism* are an example of such a reformulation. In this context earth (or other rotating massive body) is analogous to a uniformly charged rotating sphere and the Lense-Thirring result is analogous to the Larmor precession of a small spinning charged sphere in the presence of the magnetic field from the large spinning sphere. In this way the magnitude of the Lense-Thirring precession follows from the magnetic field of the charged sphere.

The magnitude and direction of the precession vector depends on moment of inertia I and rotation at rate $\mathbf{\Omega}$, and the location of the gyroscope, in this case at distance $\mathbf{R}$ from the centre of rotation. Following Stedman [112], the induced precession rate is

$$\mathbf{\Omega}_{LT} = \frac{GI}{c^2 R^3}\left[\frac{3\mathbf{R}}{R^2}\left(\mathbf{\Omega}\cdot\mathbf{R}\right) - \mathbf{\Omega}\right] \tag{5.11}$$

where c is the speed of light and G is the gravitational constant [4].

Evaluating the vector product for a gyroscope fixed at radius R_E and latitude ϕ, the magnitude of the precession vector is

$$\Omega_{LT} = \Omega_E\left(\frac{GI_E}{c^2 R_E^3}\right)\sqrt{1+3\sin^2\phi} \tag{5.12}$$

and the direction is

$$\theta_S = \tan^{-1}\left(\frac{1}{2\tan\phi}\right) \tag{5.13}$$

where θ_S is the angle of tilt from local vertical towards the south. This gives a direction outwards from the earth in the northern hemisphere and inwards in the southern hemisphere.

If we consider a gyroscope built with the plane perpendicular to the earth rotation axis (i.e. minimal sensitivity to tilt changes), and evaluate the vector product (Equation 5.11) we find the observable component of the rotation rate due to Lense-Thirring precession is

$$\Omega_{LT} = \Omega_E\left(\frac{GI_E}{c^2 R_E^3}\right)(3\sin^2\phi - 1). \tag{5.14}$$

This has a maximum at each pole, minimum at the mid latitudes and one half the maximum value in the opposite direction at the equator as shown in Figure 5.7. The maximum precession rate is $\Omega_E \times 4.6 \times 10^{-10}$. By comparison, the current best Allan

deviation achieved by a ring laser gyroscope (see Figure 2.7, G laser) is approximately 2 orders of magnitude larger.

Two approaches could be taken to measuring the Lense-Thirring precession. Either the deviation in angle of the earth rotation vector could be measured, or the deviation in magnitude.

Taking the approach of measuring a change in the angle of the rotation vector has the advantage that the measurement is independent of the rotation rate. It might possibly be implemented with multiple instruments at different orientations at the same site, with the relative orientations of each measured very precisely. The disadvantage of this however lies in the extreme orientation precision that would be required.

In a UG-3 sized instrument, an offset in the measured rotation rate of the order $\Omega_E \times 1 \times 10^{-10}$ could be caused by a change in the tilt of the laser plane of 1×10^{-10} rad. This means it would be necessary to survey the true coordinates of the reflected points to an accuracy of around 10 nm. This appears a virtually insurmountable problem. For this reason we advocate measuring the change in rotation rate directly, with an instrument built to be insensitive to changes in tilt. The measured rotation rate would be compared to the best measurements from VLBI[1] and possibly other ring laser gyros.

5.5. Thomas and de Sitter precession

The Thomas precession is a result of the kinematics of special relativity applied to circular motion. It predicts that a gyroscope transported around a circular path will acquire a precession. It was first described by Llewellyn Thomas in 1927 [113] in the context of electron spin.

The effect can be understood by considering a gyroscope transported around a circular path in flat space (for the moment we neglect gravity, or assume the transportation is arranged in some way other than an orbit). The circular path can be approximated as a polygon with a large number of sides N. In the inertial frame, the gyroscope has rotated by $\frac{2\pi}{N}$ for each segment. In the gyroscope frame however, special relativistic length contraction applies along the velocity vector $\mathbf{v}$, and not along the component

[1] Very long base-line interferometry, a technique used in radio astronomy.

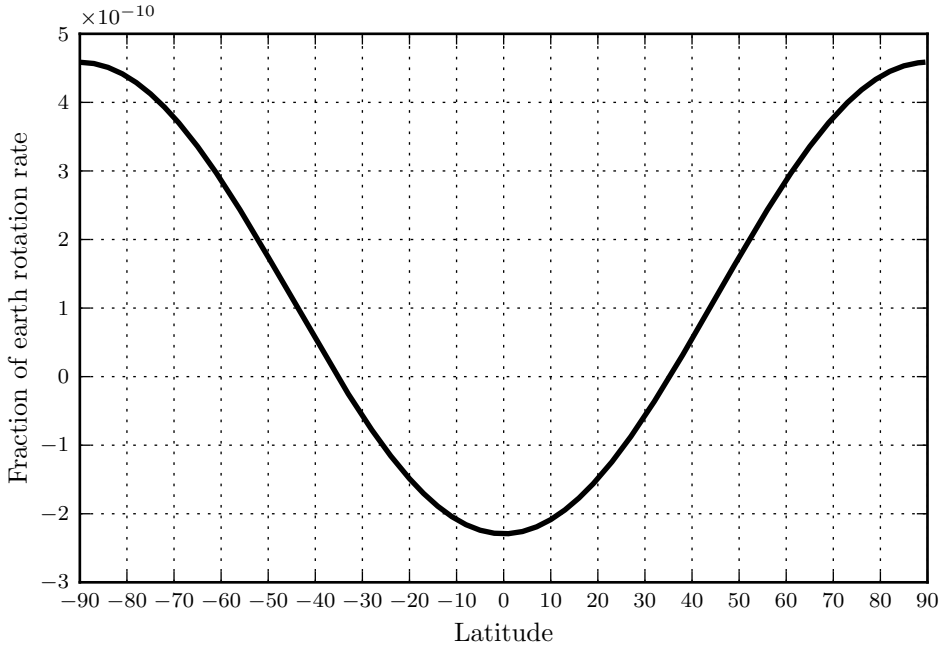


Figure 5.7. Lense-Thirring precession rate for a gyroscope perpendicular to the earth rotation axis.

directed towards the centre of rotation. After the gyro has made one circuit, the angle of rotation observed by the gyro is

$$\theta = 2\pi\gamma \tag{5.15}$$

where the Lorentz factor

$$\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \tag{5.16}$$

(see Belušević [114]). From the difference in angles in the two frames, the Thomas precession rate can be found,

$$\Omega_T = \frac{-2\pi(\gamma - 1)}{1 \text{ orbit}}. \tag{5.17}$$

Thomas precession can be described by applying only special relativity in flat space, it is a direct result of the non-commutativity of velocity additions in special relativity. The de Sitter precession however (also called the geodetic precession) incorporates the

Thomas precession in curved space as described by general relativity. While general relativistic, the key difference between this and the Lense-Thirring precession is that the latter is due to the rotation of the central mass where as the de Sitter precession applies regardless of the rotation of the central mass.

While a full treatment of de Sitter precession is quite involved, it turns out that for a gyroscope in orbit around a massive body a simple relation emerges for the magnitude of the precession (see Rinder [115]),

$$\Omega_{DS} = 3\Omega_T. \tag{5.18}$$

To the authors knowledge, the de Sitter precession caused by earth rotation as measured by an observer fixed to the earth lacks a direct theoretical treatment at present. There is however a surprisingly large precession due to the earth's rotation about the sun which can be calculated from Equation 5.18.

Substituting the equivalent numbers into Equation 5.18 (radius 1 AU, orbital period 31 557 600 s) we find the de Sitter precession has magnitude of $\Omega_E \times 4 \times 10^{-11}$, approximately an order of magnitude larger than for earth rotation and one order of magnitude smaller than the Lense-Thirring precession. However, this precession may be easier to measure. Even for a gyroscope with the plane perpendicular to the earth's rotation axis, the observed rotation precession rate should vary with time due to earth rotation and axis tilt and the ellipticity of the earth's orbit. It is generally much easier to recover a modulated signal in noise than a constant offset, even if the modulation is quite small.

In the recently completed Gravity Probe B experiment [116], four mechanical gyroscopes with axes oriented in order to measure a torque due to the Lense-Thirring and de Sitter effects were placed in a polar crossing orbit about the earth. In general the results from this experiment were below expectations due to unexpected noise. Nevertheless, they were able to confirm the de Sitter precession. The Lense-Thirring precession is smaller for their experiment and initially below the measurement noise level, though investigations to apply corrections and extract a Lense-Thirring signal continue.

The general disappointment with the noise levels in the Gravity Probe B experiment

provides a strong motivation for a renewed effort. There have recent calls [117, 118] for terrestrial measurements of the Lense-Thirring effect, because these provide stronger tests of some aspects of general relativity than some of the space-based measurements. If successful, a ring laser measurement of either effect, would be the first such terrestrial measurement.

5.6. Applications in gravity wave detectors

Since the publication of the famous 1916 paper on general relativity by Einstein [119], the direct detection of gravitational waves has remained an elusive goal. Routine observation of gravitational waves could offer a new window on the universe, complementing optical/radio and neutrino astronomy.

Throughout the history of large ring laser gyros, speculations have occasionally been made regarding the detection of rotational components of gravity waves. These effects are generally considered much too small to be detected by a ring laser gyro [4]. Nevertheless, ring laser gyros may have useful applications to the kilometer-scale Michelson interferometers that are currently under development for this task.

The basic idea underlying the very large Michelson interferometers intended for detecting gravity waves is to measure tiny relative changes in the arm lengths. There are at present five[1] notable devices of this design operating. This section focuses on the joint French-Italian VIRGO interferometer. This instrument has a 3 km arm length and is located near Pisa, Italy.

5.6.1. Background

Potential sources of gravitational radiation include periodic signals from elliptically deformed pulsars (upper limit in strain amplitude of 1.5×10^{-24} [120]), and impulse sources, for example the supernova explosions which occur in our galaxy at a rate of about 2.4 each year [121] give strain signals in the range of 10^{-24} in amplitude.

The gravitational wave detectors developed to date are usually considered first generation instruments, they represent the first instruments that have a real chance of detecting gravity waves. However, no gravity waves have been detected to date. This is not unexpected, as based on the most optimistic astrophysical estimates there are

[1] LIGO, VIRGO, CLIO, GEO 600, TAMA 300

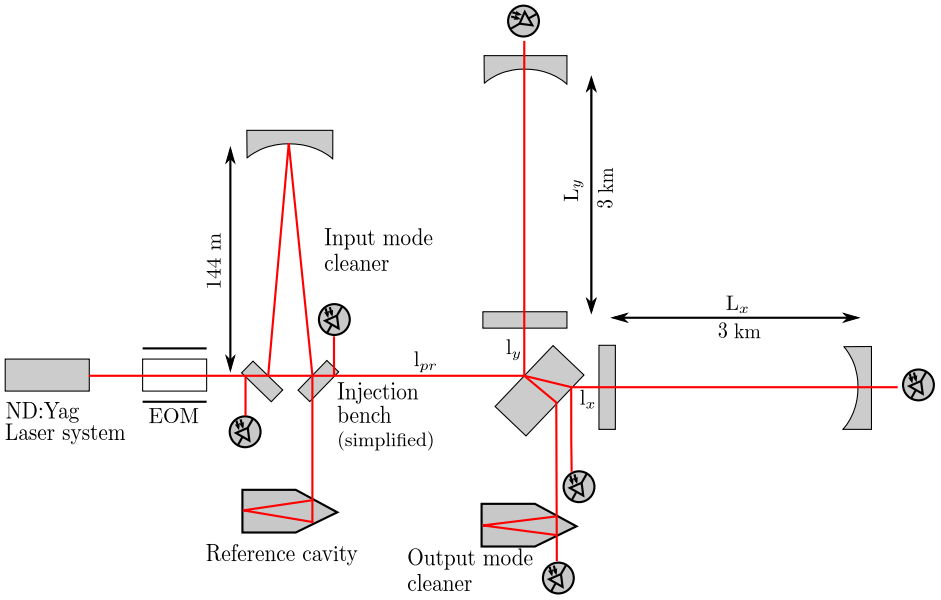


Figure 5.8. Simplified diagram of the VIRGO interferometer, showing only the basic optical arrangement and the minimal set of detectors necessary for longitudinal control.

only a few potentially detectable events per year. For this reason a major upgrade, *Advanced* VIRGO/LIGO has been proposed. The goal of this upgrade is an increase in sensitivity over the whole detection band of about one order of magnitude. A detector realising these specifications is expected to see many events every year and thus usher in the era of gravitational wave astronomy.

Before advanced VIRGO/LIGO can be realised however, new techniques and technologies must be demonstrated. As inertial sensors, ring laser gyros have unprecedented performance and no moving parts. They could form ideal sensors for controlling the alignment of interferometers such as VIRGO/LIGO [122, 123].

The basic optical arrangement of the VIRGO instrument consists of a Michelson interferometer with two orthogonal equal-length arms, very similar to that originally presented by Michelson (1887) [124]. Beyond the basic optical setup however, there are several differences. A simplified diagram of the optical arrangement is shown in Figure 5.8.

Firstly, a 20 W, 1064 nm laser system (Nd:YVO_4 locked to a 1 W Nd:Yag), followed by a 288 m perimeter triangular optical mode-cleaner cavity provides an extremely

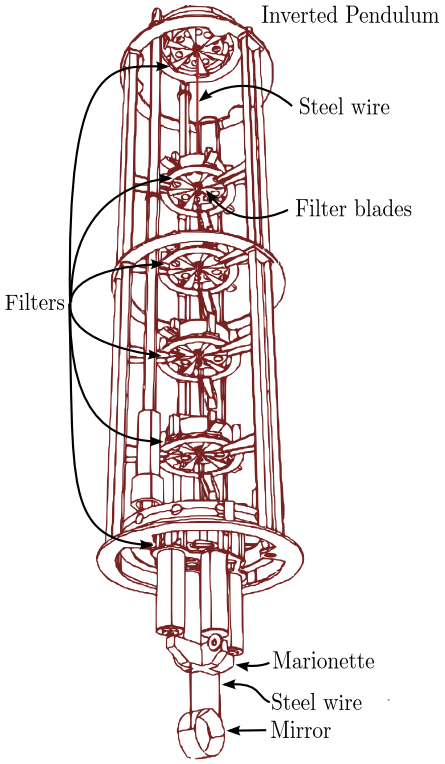


Figure 5.9. VIRGO super attenuator. Modified from a figure presented by Braccini *et al.* [125].

well stabilised light source.

Secondly, each arm contains a Fabry-Pérot 'power recycling' cavity. The interferometer is controlled to maintain a resonant condition inside these cavities (said to be the *operating point*). Overall the Fabry-Pérot cavities effectively increase the arm length by 50 times.

Thirdly, the input beam is modulated at radio frequency (6.26 MHz and 22 MHz) by electro-optic (EO) modulators. This allows a *Pound-Drever-Hall* [126] technique to be applied. This technique can be used to obtain a set of four signals (shown in Table 5.1) which in turn provide the input to a servo system controlling the mirror positions. These positions are controlled in order to maintain resonance in the Fabry-Pérot cavities, and the final output carrier is held at minimum intensity ('dark fringe' operation). The 'signal' (i.e. a gravitational wave) causes an increase in intensity of the output from zero. This configuration maximises signal-to-noise [127].

In addition to control of the various path lengths, alignment control is a major challenge. The optics must be isolated from seismic and other noise sources to an extremely high degree. This is achieved by mounting the optical elements on elaborate suspension systems called *super attenuators*. These are large cascades of compound pendula assembled in ultra high vacuum. The standard VIRGO suspension is assembled in a 10 m vacuum chamber. The first stage consists of a 7 m tall, three-legged inverted pendulum which is actively controlled in translation and yaw by mechanical accelerometers. This stage filters out the slowest (∼30 mHz) ground vibrations.

Attached to the top of the inverted pendulum is a cascade of 5 pendula which form a passive filter. The individual pendula (filters) are drum-shaped steel structures 0.8 m in diameter weighing about 100 kg. They contain a set of carefully designed triangular cantilever steel blade springs, the tips of which are connected through 1 mm diameter steel wires to an internal crossbar and cylindrical column. From this column additional steel wires support the following filter(s). The final filtering stage is called the 'marionette'. This supports the 20 kg, 350 mm diameter fused silica interferometer mirrors in a carefully designed cradle [125].

The marionette can be moved through electromagnetic actuators to change the orientation of the mirror. These actuators are controlled by an elaborate system of auxiliary alignment optics employing many optical lever systems, quadrant photodi-

Name	Length
Differential arm	$L_x - L_y$
Common arm	$L_x + L_y$
Power-recycling cavity	$l_{pr} + \frac{l_x+l_y}{2}$
Michelson length	$l_x - l_y$

Table 5.1. Longitudinal degrees of freedom which can be controlled to lock the interferometer.

odes and CCD arrays. Overall this system can control RMS alignment fluctuations to a level of 1 µrad (bandwidth 4 Hz) [128].

Lastly, an output mode cleaner is used to further enhance the signal to noise ratio.

The basic principle of the Pound-Drever-Hall technique can be illustrated by assuming the Fabry-Pérot cavity is held on resonance for an unmodulated carrier, which is stabilised to high precision. Under these conditions a large fraction of the carrier will be transmitted. A modulation frequency is chosen such that the induced sidebands do not coincide with a multiple of the cavity's FSR, and are thus mostly reflected by the cavity. Photo-detectors measuring the reflected light will detect an optical beat signal generated by the carrier and the modulation-induced sidebands. This signal has a component at the modulation frequency, and this gives a measure of the phase difference between the carrier and sidebands. Any slight change of the cavity length causes a change in the phase of the carrier and no change in the sideband. Overall this allows changes in length of the cavity to be measured to high precision and results in a movement of the measurement band to the low noise, MHz domain.

5.6.2. Tilt control

At present the control system for the inverted pendulum works only in four degrees of freedom; three translational and yaw. Due to the *cradle effect*, the linear accelerometers providing these feedback signals are fundamentally unable to distinguish between longitudinal accelerations and momentary changes in inclination. Since the earth rotation signal measured by a ring laser changes with inclination, a system of two ring laser gyros could be applied as precision tilt meters. Changes in inclination could then be uniquely determined. Such a system is expected to reduce the overall RMS

displacements of the inverted pendulum, which would in turn translate into improved uptime for the instrument.

One option for employing a ring laser in this application is to simply install the system outside of the tower. In this way the physical size and vacuum compatibility is inconsequential and the system could easily be serviced. However tilt signals being measured are only quite indirectly related to the correction signals that must be applied.

Another option would be to put the system of gyros inside the vacuum tank, on top of the inverted pendulum. This would be far better from a scientific point of view, however is more difficult in practice. In this option, size becomes an issue. A device with area less than $0.7\,\mathrm{m}^2$ is required to fit inside the system. A ring laser with reasonable performance capable of unlocking on earth rotation alone within this size constraint has not previously been demonstrated. Additional problems to be solved include the excitation of the gain medium when the whole device is under vacuum, and the long periods of operation between servicing which would be required.

In order to avoid introducing additional noise into the system, the requirements following the advanced VIRGO specification indicate a tiltmeter with sensitivity of at least $10^{-8}\,\mathrm{rad}/\sqrt{\mathrm{Hz}}$ at 10 mHz. This is comparable to an Allan deviation in earth rotation rate measurement of $1 \times 10^{-6}\Omega_E$ at 100 s. By comparison with Figure 2.7, existing instruments UG and G exceed this specification by at least an order of magnitude. Although it would be necessary to build a smaller instrument for this task, it is still likely to easily exceed the advanced VIRGO specification.

5.6.3. Alignment Control

In general the closer an inertial sensor can be installed to the interferometer mirrors, the more effective it can be. However, the closer it is to the interferometer mirrors, the more stringent the noise requirements. To this end, we consider the application of a system of ring laser gyros used for direct control of the interferometer mirror. Such a system could extend the scope of the alignment control to all six degrees of freedom. Alignment of the interferometer could be maintained to higher accuracy and the low frequency response of the interferometer improved. In addition, a reduction in various non-linear effects over the whole measurement range could also be expected.

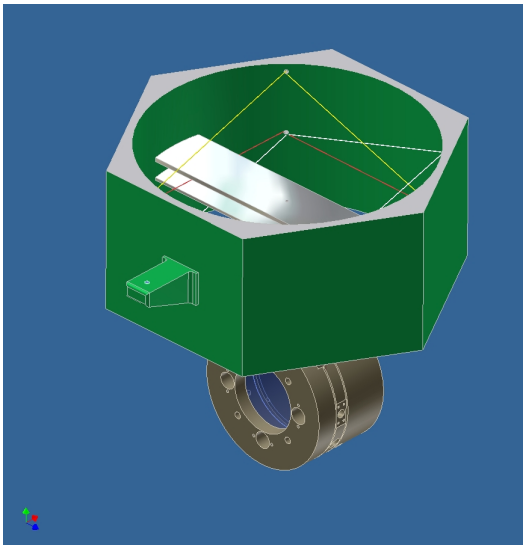


Figure 5.10. Example configuration of a 3-axis ring laser system in a modified marionette

The basic idea is to install a 3-axis system of ring laser gyros in one of the pendula. The signals could then be used in a feedback loop to control the magnetic actuators of the preceding pendula. Figure 5.10 shows an example rendering of how this might be implemented in a modified marionette. Note that it is not strictly necessary to arrange the lasers orthogonal to each other, given the size constraints a compromise here is necessary.

A preliminary analysis of the the noise such a system would inject into the alignment control has been performed by Di Virgilio[1]. This was done with a simplified mathematical model of a '4+1' (four filters plus the final stage as depicted in Figure 5.11) VIRGO suspension system[2]. A ring laser gyro with the relatively ambitious (especially considering that to be physically compatible with existing suspensions the gyros must have a maximum side length of around 60 cm.) resolution of $10 \times 10^{-11}\,\mathrm{rad\,s^{-1}}$ was assumed.

It was found that with the slight compromise of placing the system of gyros one stage above the final mirror stage, satisfactory performance could be obtained.

[1] Unpublished

[2] The parameters (spring stiffnesses, moments of inertia etc) used in the model were manually tuned in order that the suspension response match the measured VIRGO transfer functions [129].

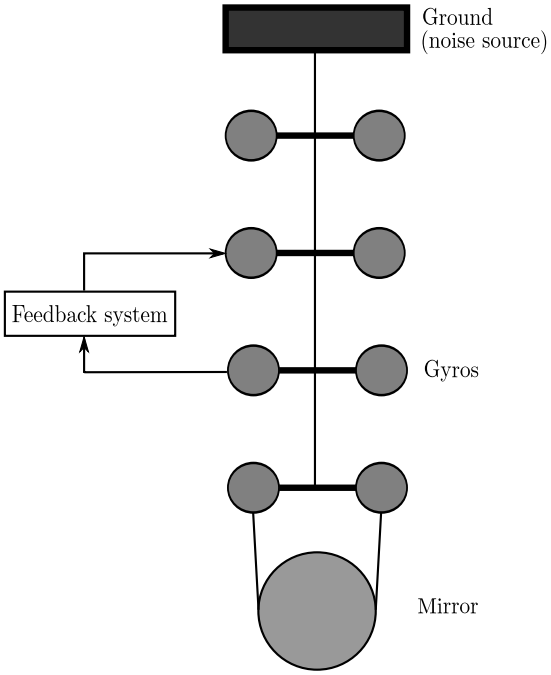


Figure 5.11. Simple suspension model with ring laser gyro feedback.

6. Data Acquisition and Storage

The goals of this chapter are to outline the data life-cycle typically applied in ring laser research (raw data acquisition, though validation and analysis), then identify the major challenges and inefficiencies. An underlying theme is the need for rigorous standardisation in data acquisition and storage. To this end we present an XML data specification which has been developed for storing ring laser data. We also demonstrate techniques that have been used in data acquisition in other parts of this thesis, and which have been encapsulated into a 'tool box' of routines which has been developed for data acquisition and analysis.

6.1. Background

Presently there are five[1] sites around the world known to the author, operating a total of eight ring laser gyroscopes. Some are intended to be operated in a minimal disturbance 'observatory' mode, while others are intended as experimental instruments and are regularly modified.

Data has traditionally been acquired with a single desktop computer system running LabVIEW software at each site. The large number of signals which must sometimes be acquired and lack of separation of concerns has lead to maintainability and efficiency problems. An example is the acquisition and storage of slowly varying signals at unnecessarily high speed due to mixing different discrete instruments in one acquisition system.

Data has traditionally been stored in a variety of application specific text and binary formats. There has been minimal standardisation and quality control in the data formats. Analysis has been a mostly manual process involving searching data stored on

[1] Cashmere Cavern (NZ), Rutherford Building (NZ), Wettzell (Germany), Pisa (Italy), Piñon Flat (CA, USA).

optical media, then decoding and eliminating erroneous data with a variety of tools. This has been found to encourage a scientific data life-cycle involving many discrete tools and make reanalysis of data from primary sources difficult (Figure 6.1). The motivation for moving to a system with more standardisation is to encourage a more efficient life-cycle (Figure 6.2). Standardisation encourages re-use of components and allows leveraging modern tools for efficient assembly and manipulation of data across multiple instruments.

6.2. Requirements

Starting from the point of view of obtaining science results as efficiently as possible, we can work backwards to define the requirements of a complete system for data acquisition and storage of ring laser data.

1. Data should be easily accessible, preferably human-readable and easily transformable to formats suitable for use with a variety of analysis tools used in related fields. The notation used should conform to the notation conventions used in rotational seismology and other related fields.
2. There should be an objective measure of data quality at each stage of the cycle from acquisition through to final analysis. To this end, a level of redundancy and assurance against data corruption are necessary.
3. It should be possible to run complex queries across data, for example it should be possible to program a query such as (in natural language) "find all Sagnac waveforms recorded on any instrument, during an earthquake with Richter magnitude between a particular range and only on days with minimal microseismic activity."
4. It should be possible to store many different types of data, not limited to direct time-series data. Frequency domain measurements and images should also be accommodated for.
5. Measurements should have accurate timing information, allowing side by side comparison of data obtained from various systems. While true real-time acquisition may not be necessary for all systems, provision should be made for this mode of operation.

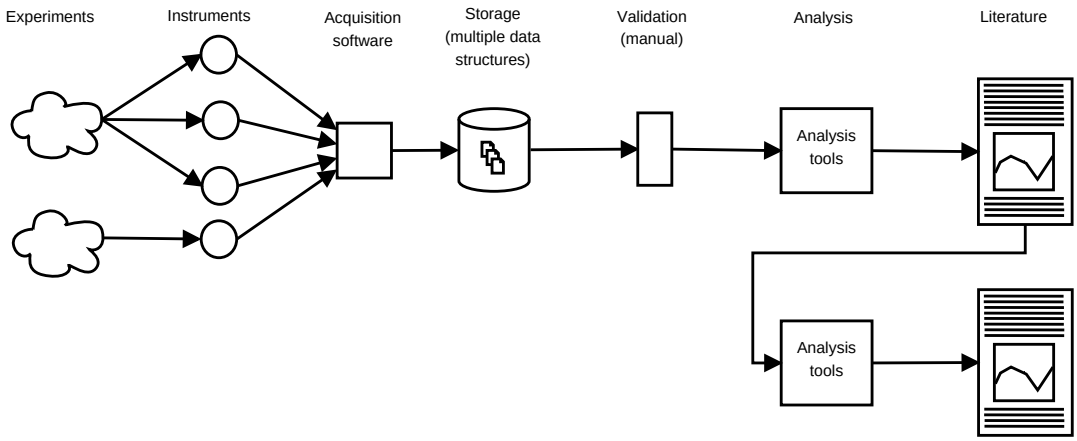


Figure 6.1. Lack of a standardisation leads to a data life-cycle with minimal opportunity to re-use both software components and raw data forming the basis of published literature.

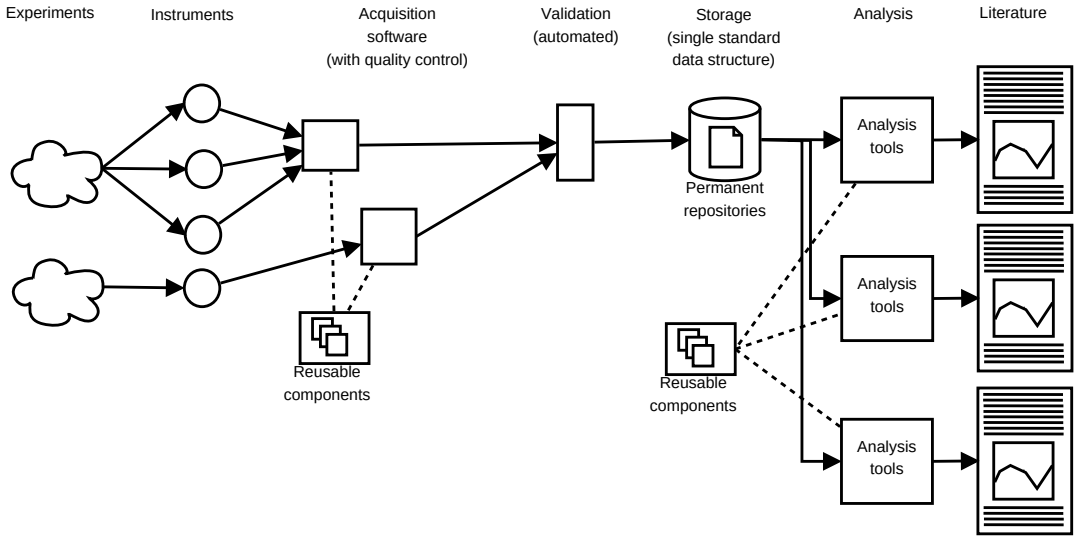


Figure 6.2. A data life-cycle where standardisation and separation of concerns encourages data sharing and efficient analysis.

6. Software components should be reusable across multiple platforms.

7. A basic level of communication is necessary between all data acquisition systems. For example it should be possible for one system to report the occurrence of an earthquake to another with minimal delay so that appropriate action can be taken.

8. Raw data acquisition rates should be fast enough so as to not discard potentially valuable data, for example Sagnac waveforms must (at a minimum) be obtained at a sufficient rate to mitigate aliasing. The existence of the microseismic background implies the Sagnac frequencies should be measured at a rate of around 5 Hz. Achieving good accuracy in these measurements requires sampling the underlying Sagnac waveform well in excess of the Nyquest rate. For example, with the UG laser (Sagnac frequency 1.5 kHz), 10 kHz sampling is appropriate.

9. It should be possible to save raw data at full acquisition rate in the event of a seismic event for later analysis.

10. Acquired data should be sufficiently compact so as to allow it to be easily sent across available computer networks, stored and indexed.

6.3. Design Principle: Separation of Concerns

Separation of concerns is well established as an important principle in software engineering. It is achieved by encapsulating the logic required to perform individual tasks so that each task can operate independently. When data needs to be passed between tasks, it will in turn be encapsulated into well defined data structures, again specific to particular concerns. Historically we see that employment of this principle leads to robustness, maintainability, and re-usability [130]. Components designed with this in mind are said to be *cohesive*.

Examples of separate concerns in our application are the individual ring lasers, data storage and analysis systems. The specific requirements of acquisition from one instrument (and any problems which occur) should not be able to affect the acquisition of data on other instruments.

Another example is the separation of acquisition software into parallel tasks for ac-

quiring raw data (which must be done at high speed), processing raw data, output of processed data and monitoring. Delays in processing of raw data must not affect the acquisition or responsiveness of the user interface.

6.4. Design Principle: Anticipation of Change

It is not reasonable to assume that the best solution will always be implemented initially, and even if this were the case it is certain that the scope of the experiment will change in the future and technologies will become outdated. The approach of designing a single large acquisition system to satisfy all the design requirements from the outset is almost certain to lead to an unmaintainable system in the long term. Instead the approach that should be taken is to design a series of standard, documented components from which a data acquisition, instrument control or storage system can be assembled in minimal time.

Strong coupling is a major obstacle to change. If two components are strongly coupled then it is likely that changes to one will cause problems for the other. For this reason, platform/LabVIEW specific methods of storing data or passing data between systems should be avoided.

In software engineering, *cohesion* is a measure of how strongly related or focused the responsibilities of an individual module are. Cohesive components are easier to reuse when requirements change. Conversely, when several tasks have been rolled up into one component it is likely that the component will need to be split up or extensively modified when changes are necessary.

6.5. XML Data Format

Central to the idealised data life-cycle illustrated in Figure 6.2 is the use of a standardised hierarchal data format. Binary formats such as HDF were considered along with text based XML (Extensible Markup Language)[1] and binary XML formats[2]. Standard text based XML was chosen due to:

[1] XML is a text-based open standard defined by the W3C for encoding documents and other data structures electronically.

[2] Various binary XML formats are coming into favour in some scientific applications where very large data sets need to be processed at high speed. No dominant standard has emerged at the time of writing.

1. The inherent advantages of human-readability.
2. The ability to automatically validate data against a formal specification [131]. Validation after acquisition ensures that any inconsistencies are swiftly discovered (for example, neglecting to specify the units of a measurement) and the resulting data can be stored with a high level of confidence that it is not ambiguous.
3. The ability to make rich queries across large amounts of data using XQuery [132] and XPath [133].
4. The existence of tools such as XSLT and XQuery which can be used to transform data into other formats used in related fields. Examples include the SEED and QuakeML [134] formats used in seismology and XSIL [135] which is used in gravity wave detection.
5. The increasing popularity of XML based formats across a wide range of scientific applications [136].

A ring laser XML data format has been developed taking into account our specific requirements in order to ensure that the data gathered is not ambiguous. Additionally, the notation conventions used in other related formats have been taken into account.

The ring laser XML data format is formally defined by a RELAXNG schema [131]. Because the specification is machine readable, data can be automatically validated against the specification to ensure it is not ambiguous and issue a warning if some aspect of the data (e.g. units, sensor type etc) have not been defined.

A 'tool box' of LabVIEW VIs and associated scripts has been developed for ring laser data acquisition. These are documented in Section B.4. This includes VIs (Section B.4.6) and associated scripts (Section B.2.2, Section B.2.3) for generating XML data following this specification.

The basic structure of the ring laser XML data consists of `<instrument>` elements containing `<sensor>` elements. Different `<sensor>` types can be selected by attributes to the `<sensor>` tag, including time domain sensors, frequency domain and image sensors. The `<sensor>` elements contain `<measurement>` elements. The contents of `<measurement>` elements depend on the attributes of the parent sensor (e.g.

value and optionally uncertainty for a numeric sensor). All measurements must contain one or more `<time>` elements specifying the time value and format. Individual measurements can be encapsulated in `<event>` elements to indicate seismic events and a corresponding `<eventinfo>` element can contain additional information about the event.

Appendix B.3 contains the definition of an XML grammar for storing ring laser data and is marked up in RelaxNG format. The following sections give a summary of this format in natural language. The data types used in the schema are as defined by the W3C XML Schema data type definitions (2001)[1].

6.5.1. Root node

All XML documents represent a tree structure, and as such are required to have exactly one root element. In our case this is the `<rldata>` element.

This element can contain only `<instrument>` and `<eventinfo>` elements. Any number of these elements can occur in any order.

6.5.2. Instruments

The `<instrument>` is intended to represent a physical device, which in turn has a number of sensors associated with it. For example a ring laser or seismometer. In some cases it may be necessary to be relatively abstract with the definition of an instrument, for example it is probably most convenient to encapsulate a group of temperature sensors controlled by the same hardware as a single instrument.

Instruments contain only `<sensor>` elements.

The instrument elements requires a compulsory `name` attribute. Instrument names should be short and consistent. It is recommended that the names be kept to 5 characters, or at least that the first 5 characters are unique, in order to maintain compatibility with the 5-byte station identifier codes used in the SEED format. For convenience it is also recommended that these attributes are lower case and do not contain white space, though this is not a formal requirement.

Instruments allow a number of additional text attributes which are optional but

[1] http://www.w3.org/2001/XMLSchema-datatypes

strongly recommended (in particular the first three). The presently available attributes are shown in Table 6.1.

Attribute	Example value
location	cashmere
latitude	-43.5746
longitude	172.6232
configuration	singlemode
Ne20_pressure	0.675 mbar
Ne22_pressure	6.75 mbar
He_pressure	9.25 mbar

Table 6.1. Optional instrument attributes.

6.5.3. Sensors

The <sensor> is intended to represent a physical or virtual[1] sensor associated with the parent instrument.

Sensor elements can contain any number of <measurement> elements. They may optionally be grouped together in <event> structures.

All sensor elements must have a name element containing a text string such as 'NE camera'. They may optionally have a type attribute describing the sensor, for example 'sony'.

There are at present[2] three types of sensors available. The type of sensor is selected through two compulsory attributes; domain which may be either 'time' or 'frequency' and datatype which may be either 'numeric' or 'image'.

Numeric sensor types may also have a bindata attribute with possible values of float32, float64 or float128. See <binaryseries> for more information about this option.

[1] For example, a program which calculates a power spectra from existing data would be considered a virtual sensor.

[2] The design of the schema is such to allow this to be easily expanded.

6.5.4. Measurements

The `<measurement>` element is intended to encapsulate all information read from the parent sensor.

All measurements must contain one or more `<time>` elements, representing the time the measurement was taken. Further specific requirements depend on the type of sensor selected, and are detailed below.

6.5.5. Numeric sensor types

Selecting numeric as a data type requires that each following `<measurement>` element contains exactly one `<value>` element (containing either only decimal data or a `<binaryseries>`). Optionally, numeric sensors may contain a decimal `<error>` element, representing the absolute error[1] on the measured value. Additionally, the following requirements on the sensor attributes are selected:

1. May optionally have a decimal `interval` attribute defining the total interval of time in seconds over which the individual measurements represent the physical data. Should a `interval` attribute not be given, it can be assumed that the following measurements can be considered instantaneous.

2. Must have a `units` attribute, an arbitrary string describing the units used in child `<value>` elements. The standard SI units and prefixes are strongly recommended. Units raised to powers other than one should be written using the `^` notation. The `/` character is also acceptable, e.g. `m/s^2`. Dimensionless quantities should use the word `ratio`.

```
1 <rldata>
2     <instrument name='ug3' location='cashmere'>
3         <sensor name='sagnac' domain='time' datatype='numeric' type='counter'
              units='Hz'>
4             <measurement>
5                 <time format='UTC'>1072950759</time>
6                 <value>1532.334322</value>
7             </measurement>
8             <measurement>
9                 <time format='UTC'>1072950760</time>
```

[1] I.e. the error in the units of the measurement rather than the fractional error. There is no formal requirements as to the statistical properties of the error, though usually 1σ would be used.

```
10                  <value>1532.873243</value>
11              </measurement>
12          </sensor>
13      </instrument>
14  </rldata>
```

Listing 6.1 An example ring laser XML file (valid according Appendix B.3) showing an instrument with two measurements from a numeric sensor.

6.5.6. Image sensors

Selecting image as a data type requires that any following `<measurement>` child elements contain a `<imagedata>` element. Additionally the `<sensor>` element must have the following attributes:

1. `encoding`, a text string describing the character encoding of the data in child `imagedata` elements. For example, `base64`.
2. `format`, a text string describing the format of the encoded image, for example `TIFF` or `JPEG`.
3. The optional `interval` attribute is also available (see Section 6.5.5).

Image sensors are only available in the time domain.

```
1  <rldata>
2      <instrument name='ug3' location='cashmere'>
3          <sensor name='ne_camera' type='sony' domain='time' datatype='image'
               encoding='base64' format='TIFF'>
4              <measurement>
5                  <time format='UTC'>1072950762</time>
6                  <imagedata>TWFuIGlzIGRpc8Rbgd43agNZ4dFb24IddfCBieSB0aGlz</
                       imagedata>
7              </measurement>
8          </sensor>
9      </instrument>
10 </rldata>
```

Listing 6.2 An example ring laser XML file (valid according to Appendix B.3) showing an instrument with a measurement from an image sensor.

6.5.7. Frequency domain sensors

Selecting `frequency` as the domain of a sensor gives the enclosed measurements the same requirements as for a numeric sensor (Section 6.5.5), with the following additional constraints:

1. The measurement must contain a decimal `<frequency>` element.
2. The sensor element `freq_units` must be defined, this is an arbitrary string representing the units that the `frequency` element of the sensors child measurement elements must have.
3. Optionally the sensor may have a decimal `freq_interval` attribute, defining (in the units selected above) the frequency interval over which the measured values corresponds.

```
<rldata>
    <instrument name='ug3' location='cashmere'>
        <sensor name='powerspect' domain='frequency' datatype='numeric' units=
            'dB' freq_units='Hz' freq_interval='0.1'>
            <measurement>
                <time format='UTC'>1072950762</time>
                <frequency>100</frequency>
                <value>0.15</value>
            </measurement>
            <measurement>
                <time format='UTC'>1072950762</time>
                <value>0.18</value>
                <frequency>101</frequency>
                <error>0.02</error>
            </measurement>
            <measurement>
                <time format='UTC'>1072950762</time>
                <frequency>100</frequency>
                <value>0.15</value>
            </measurement>
        </sensor>
    </instrument>
</rldata>
```

Listing 6.3 An example ring laser XML file (valid according to Appendix B.3) showing an instrument with a measurement from a frequency domain sensor.

6.5.8. Representing time

The `<time>` element is used throughout the schema for representing timing information. It must contain only decimal data or a `<binaryseries>` element. Because of the importance of accurate timing, each time element has a compulsory `format` attribute. It is possible to define one or more `<time>` elements with different formats. The `format` attribute must take either one of two values:

1. `'UTC'` — indicates that the data represents UTC time encoded as Unix time stamp[1]. It is useful as a general purpose time format, easily supported by almost all modern programming languages which typically include the functionality to represent it in many different formats. Normally this time format will be provided by the computer clock of the data acquisition system. Note that the inclusion of leap seconds in UTC time can make calculations using this time format ambiguous across leap second boundaries.

2. `'MJD'` — indicates that the data represents the time in the Modified Julian Day format[2], a continuous time format. If provided in addition to UTC time, it is assumed to be more accurate, typically being derived directly from a GPS derived reference rather than from the clock of the data acquisition computer.

   Note that to obtain maximum precision in analysis, it is necessary to ensure that the data is loaded into the analysis program as an 80 bit (or larger) extended precision floating point value[3], since a regular 64 bit double precision float does not provide for μs resolution. Also note that the minimum requirement for an XML parser to be compliant with the W3C XML data types does not provide sufficient precision to fully represent sufficient numeric characters with the decimal data type, which could cause data to fail to validate. However in practice the Libxml2 library used for validation in this work does provide for a sufficient number of characters.

[1] Number of seconds elapsed since 00:00:00 UTC on January 1, 1970.

[2] As defined by the Smithsonian Astrophysical Observatory to be $JD - 2,400,000.5$ where JD is the Julian Day count starting 4713-01-01 12:00 BCE.

[3] In C, a long-double type.

```
<rldata>
    <instrument name='ug3' location='cashmere'>
        <sensor name='sagnac' domain='time' datatype='numeric' type='counter'
            units='Hz'>
            <measurement>
                <time format='UTC'>1072950759</time>
                <time format='MJD'>53005.41157407407332</time>
                <value>1532.334322</value>
            </measurement>
        </sensor>
    </instrument>
</rldata>
```

Listing 6.4 Example ring laser XML file (valid according appendix B.3) showing a measurement with multiple time formats.

6.5.9. Event information

Sometimes it is necessary to record when particular events occur that are relevant to the data gathered. For example, an earthquake occurring. These can be defined by `<eventinfo>` elements, placed in the the `<rldata>` element.

The `<eventinfo>` elements must have a `name` attribute, which can contain an arbitrary string. This string must match the corresponding `<event>` element(s) (see below). They must have one or more `<time>` elements to indicate the start time of the event. They must also have one `type` element which contains text indicating the type of event. At present the only available event type is `quake`.

If `quake` is selected as an event type, then the decimal elements specified in Table 6.2 (inherited directly from the QuakeML specification) are made available. The only differences from the QuakeML specification (OriginLocation element) is that the `<focalTime>` element is made redundant by the `<time>` element and the `referenceSystemID` is not included. Magnitude and reference number (as used by GeoNet) are also made available.

6.5.10. Applying events to measurements

It is often useful to be able to easily find measurements from various sensors that are associated with a particular event. This is done by encapsulating `<measurement>`

Element name	Data/Units
`latitude`	decimal degrees (double)
`longitude`	decimal degrees (double)
`horizontalUncertainty`	km (double)
`depth`	km (double)
`magnitude`	Richter magnitude (double)
`geonet_id`	Reference used by GeoNet (text)

Table 6.2. Optional elements available for quake event types (inherited from QuakeML).

elements inside an `<event>` element. This element has the required attribute `name` which must take the value of an event identifier as given in an `<eventinfo>` element elsewhere. Note that the file will still validate if no matching event information is present, as it may be the case that the event identifier is defined in a separate file (to be merged later) or there is no information for this particular event.

```
<rldata>
    <instrument name='ug3' location='cashmere'>
        <sensor name='sagnac' domain='time' datatype='numeric' type='counter'
            units='Hz'>
            <event name='eq35344345'>
                <measurement>
                    <time format='MJD'>53005.41158653453444</time>
                    <time format='UTC'>1072950760</time>
                    <value>975.334322</value>
                </measurement>
                <measurement>
                    <time format='MJD'>53005.411595353453344</time>
                    <time format='UTC'>1072950761</time>
                    <value>975.334322</value>
                </measurement>
            </event>
            <measurement>
                <time format='MJD'>53005.411602343323433</time>
                <time format='UTC'>1072950762</time>
                <value>975.334322</value>
            </measurement>
        </sensor>
    </instrument>
```

```
23      <eventinfo name='eq35344345'>
24          <time format='UTC'>1072950764</time>
25          <type>quake</type>
26          <magnitude>3.2</magnitude>
27          <distance>2566</distance>
28      </eventinfo>
29  </rldata>
```

Listing 6.5 An example ring laser XML file (valid according appendix B.3) showing the encapsulation of measurements with an earthquake event and the associated event information.

6.5.11. Binary series representation

For sensors recording a large number of measurements as part of a fast time series it is inconvenient to store all values in individual `<measurement>` elements as this will result in large file sizes and computationally intensive parsing and indexing.

To solve this problem an alternative representation is available, where `<binaryseries>` elements containing the binary equivalent of a series of numerical values are allowed in the place of data in `<time>` and `<value>` elements. A `<binaryseries>` element must have a `format` attribute describing how the binary data is formatted. At present the only available format is `npy`, which indicates the data in in the NumPy 'npy' data format (documented in [137]) and represented in base64 (for compatibility with XML). A Python program to transform data to and from this representation is provided.

As this representation does mean that the data is not human readable, because of this data is intended to be transformed into this format only prior to storing. This allows the data files to be assembled in a consistent way. When assembling the data files a `bindata` attribute can be added to the corresponding sensor. This indicates that a transformation to the alternate format can be applied prior to storing. The `bindata` attribute specifies the desired binary representation (see `<sensor>`).

The `<binaryseries>` elements must have an `encoding` attribute and a `format` attribute. Currently only `base64` encoding and `npy` (the array format used by the numpy and scipy set of tools) is supported by associated software, although both attributes are not restricted to this data type. If encoded in base64, a penalty of about 37% in storage efficiency is encountered, however this is mostly mitigated if

the XML file is later compressed.

Statistical information, in the form of `max`, `min`, `mean` and `stdev` attributes (which must contain decimal data) are optional but highly recommended as they can be used when querying data sets.

```
<rldata>
    <instrument name="EARSS">
        <sensor bindata="float32" datatype="numeric" domain="time" name="X
            Filtered" units="m/s">
            <event name="eq1267225456">
                <measurement>
                    <value>
                        <binaryseries encoding="base64" format="npy" max="1.2e
                            -05" mean="-5.47945205479453 07e-09" min="-1.1e-05"
                             std="6.574752256616134e-06">
                            k05VTVBZAQBGAHsnZGVzY3I...</binaryseries>
                    </value>
                    <time format="UTC">
                        <binaryseries encoding="base64" format="npy" max="
                            1267225461.6826" mean="1267225458.7626004" min="
                            1267225455.8426001" std="1.6858825577273926">
                            Je3RTi0kHr4l7dFOLSQ...</binaryseries>
                    </time>
                </measurement>
            </event>
    </sensor>
</rldata>
```

Listing 6.6 An example ring laser XML file (valid according appendix B.3) showing measurements in the binaryseries representation.

6.6. Data Acquisition Techniques

6.6.1. Timing

Accurate timing, both between systems at the same site and with respect to UTC time is important; especially for analysis of seismic phenomena. For example a absolute timing error of 1 s corresponds to an error in determining the distance to the epicentre of up to 8 km.

Earthquake motions have spectral components up to 3 Hz. If measurements from

multiple instruments are to be compared with time-frequency analysis techniques [138, 139] then relative timing between the instruments should be accurate to significantly better than the Nyquist sampling period if synchronisation artefacts are to be avoided.

Synchronisation of clocks relative to UTC is done with NTP4 which is used to establish a phase-locked-loop between the OS kernel[1] and reference clock. At the Canterbury and Wettzell sites we have local access to Stratum-1 time servers to which a host machine in each site is synchronised. This provides an accuracy of better than 300 µs relative to UTC. Each data acquisition machine is then locked to this host with similar accuracy. While the 2-step synchronisation reduces accuracy somewhat it ensures all the machines at each site will remain in synchronisation should the link to the Stratum-1 server fail.

The most convenient time system to use is UTC time, encoded as a UNIX time stamp. This is due to the large number of existing tools for manipulating and synchronising time encoded this way. However there exists some ambiguity with leap-second transitions and looking to the future it will be desirable to have multiple redundant time systems. For this reason all systems allow measurements to be associated with time stamps in both UTC and MDJ (modified Julian date) format.

6.6.2. Phase tracking

A commonly used method of monitoring the Sagnac signal is to accumulate the phase differences between the signal and a precision GPS-locked reference generator set to the nominal Sagnac frequency f_R. If multiple Sagnac signals from different corners of the same laser are tracked against a common reference, then the relative differences in phase can be used to find the changes in length of each side of the laser. This in turn can be used to correct for changes in the geometry in accordance with Equation 2.3.

Both the signal waveform $S(n)$ and reference $R(n)$ are acquired in batches of N samples at a rate of f_a. These are used with pre-calculated kernels

$$K_1 = H\left(\sin\left(2\pi f_R \frac{n}{f_a}\right)\right) \tag{6.1}$$

[1]The requirement for accurate timekeeping precludes the use of platforms not implementing NTP4, for example MS Windows which can only maintain time synchronisation in the 1 s to 2 s range [http://support.microsoft.com/kb/939322].

and

$$K_2 = H\left(\cos\left(2\pi f_R \frac{n}{f_a}\right)\right) \tag{6.2}$$

where $n = 0 \ldots N$, to obtain the phase difference

$$\Delta\phi = \frac{1}{2\pi}\left[\tan^{-1}\left(\frac{\sum_{n=0}^{n=N} K_1(n)S(n)}{\sum_{n=0}^{n=N} K_2(n)S(n)}\right) - \tan^{-1}\left(\frac{\sum_{n=0}^{n=N} K_1(n)R(n)}{\sum_{n=0}^{n=N} K_2(n)R(n)}\right)\right] \tag{6.3}$$

which is accumulated for each block. H is the Hanning window function. If the sample blocks consist of a small number of cycles then for best results the block length should be chosen so that the associated kernels contain an integer number of cycles.

The Sagnac frequency can be obtained for any period Δt from the change in accumulated phase,

$$f_s = \frac{\Delta\phi}{\Delta t} + f_R \tag{6.4}$$

and the amplitude of the Sagnac signal A can also be found conveniently with

$$A = \frac{4}{N}\sqrt{\left(\sum_{n=0}^{n=N} K_1(n)S(n)\right)^2 + \left(\sum_{n=0}^{n=N} K_2(n)S(n)\right)^2}. \tag{6.5}$$

Also of interest is the contrast ratio of the Sagnac signals (see Section 4.3)

$$C = \frac{A}{\bar{S}} \tag{6.6}$$

which is a function of beam combiner alignment and can be used to infer the angle which the beams are reflected through at each corner.

The method of phase tracking provides very accurate measurements over all time scales that the input is 'well behaved'; in which case the accuracy of frequency estimates are comparable to (or better than, depending on the noise model) direct counting. However should the frequency vary widely for example due to a seismic event, it is possible for the phase tracking to miss integer cycles which badly degrades the direct frequency estimates. Note however that the accuracy can be recovered manually by removing the integer phase step.

Phase tracking is best implemented in a class as multiple phase tracker objects often need to be instantiated in one program. This allows arrays of *phasetracker* objects to be constructed and operated on in parallel, resulting in good performing and concise code.

6.6.3. Auto-regressive frequency tracking

The 2nd order autoregressive or AR(2) technique for frequency tracking is based on a two-parameter (a_1, a_2) model describing a noise process ϵ in terms of the amplitude y of two neighbouring samples in a time series,

$$\epsilon(t) = y(t) + a_1 y(t-1) + a_2 y(t-2). \tag{6.7}$$

It can be shown that the parameters of such a model can be chosen so that the time series exhibits pseudo-periodic behaviour. The parameters of the model are estimated continuously using the Yule-Walker estimator,

$$\beta = \begin{bmatrix} C_0 & C_1 \\ C_1 & C_0 \end{bmatrix}^{-1} \begin{bmatrix} C_1 \\ C_2 \end{bmatrix} \tag{6.8}$$

where

$$C_j = T^{-1} \sum_{t=j}^{T-1} y(t) y(t-j) \tag{6.9}$$

and where T is the sample period. The frequency f can be found from

$$f = \cos^{-1}\left(-\frac{\beta_1}{2\sqrt{\beta_2}}\right). \tag{6.10}$$

The AR(2) technique is usually applied to a band pass filtered signal. The main advantages of the AR(2) technique is that it is very robust with respect to changes in frequency, and frequency estimates can be made 'on line' at a high rate. It should be well suited to dynamic situations such as seismic events[1] [2, 140]. One disadvantage is the susceptibility to noise. It can be shown to be consistent only under noise-free conditions [141]. In practice it will produce a convincing series of frequency estimates with only random noise as the input. Care must be taken to ensure the input signal is valid before processing with AR(2).

Figure 6.3 and Figure 6.4 show the results of numerical tests of the susceptibility of the AR(2) estimator to synthesised noise and comparison to two other estimators available in LabVIEW. The results show that the AR(2) performance scales poorly with noise amplitude. Gaussian, flat and 1/f noise profiles were synthesised with similar results.

[1] This may not be the case in practice if filtering has been applied prior to the AR(2) estimation.

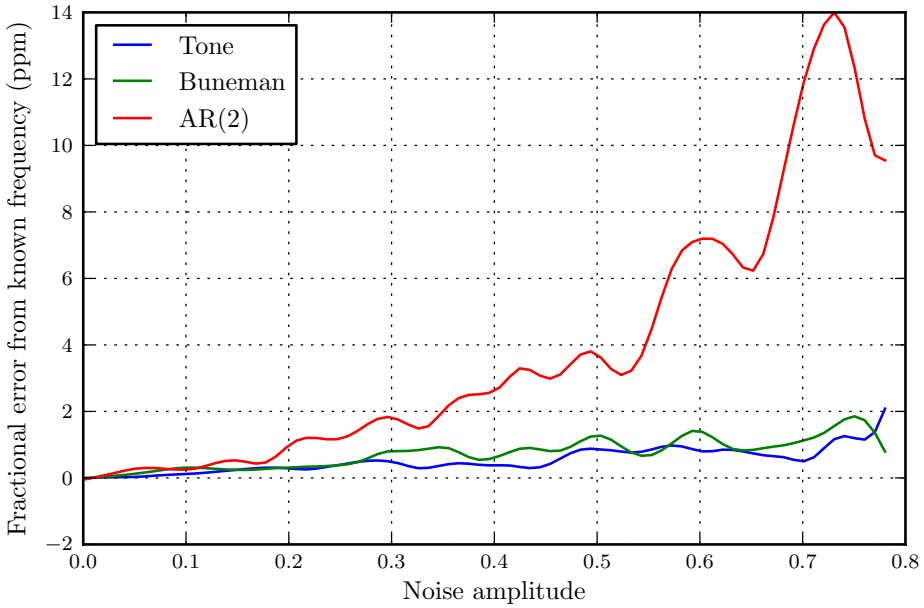


Figure 6.3. Scaling of the accuracy of frequency estimation with increasing Gaussian noise amplitude.

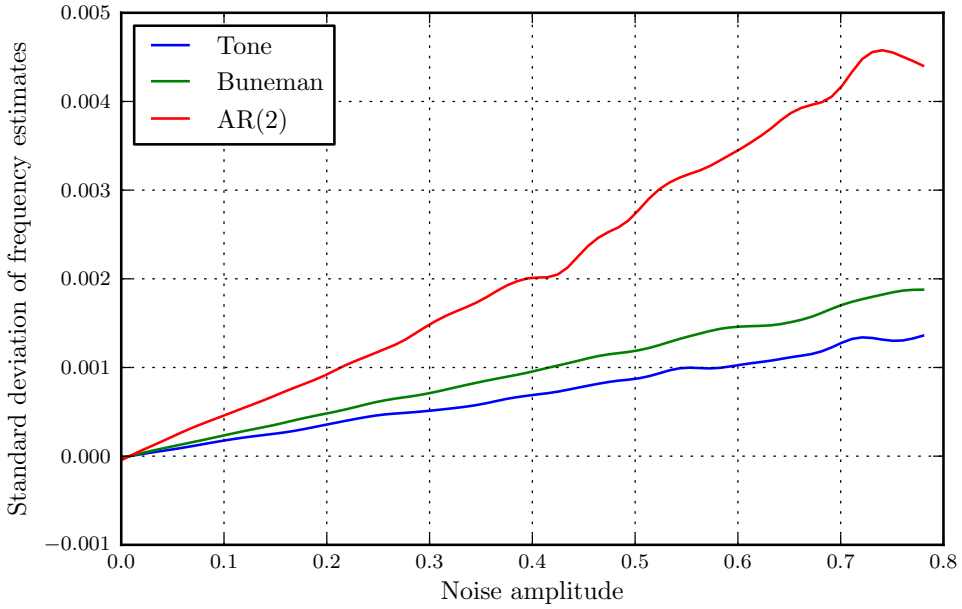


Figure 6.4. Scaling of the precision of frequency estimation with increasing Gaussian noise amplitude.

6.6.4. Counting

Direct counting of the period of the Sagnac signal is generally accepted as the most accurate estimator of Sagnac frequency over very long time periods. It can be achieved with hardware counters available on most data acquisition boards. A high frequency (usually 10 MHz GPS-locked) reference signal f_R and a Sagnac signal converted to TTL are required. The reference signal is gated by the Sagnac signal and a large number c_s of Sagnac cycles are counted. The average period of the Sagnac signal can then be obtained to a constant accuracy of one reference clock cycle over the whole measurement period t (typically around a half hour). The relative error introduced due to the stability S_R of the reference clock (typically 10^{-12}) over this period is negligible. Overall the Sagnac frequency can be obtained to an accuracy of

$$\Delta f_s = \frac{|f_s|}{t}\left(\frac{1}{f_R} + S_R t\right) \tag{6.11}$$

which is typically around 0.1 µHz. One disadvantage with this scheme is that the spacing in time of the measurements will vary with the frequency.

6.6.5. Scanning Fabry-Pérot Acquisition

The acquisition of high quality spectra from a scanning Fabry-Pérot interferometer is a task which has been necessary at several stages in this work.

Usually the output of the Fabry-Pérot will be detected by a photomultiplier and often will be quite weak. Where high Q is required it is necessary to take quite some care to optimise the alignment and obtain the narrowest possible input beam. If the beam has some astigmatism then for best results this should be corrected by an appropriate arrangement of angled spherical lenses or cylindrical lenses.

If acquisition at reasonably high speed is required (i.e. when the mode structure is changing rapidly; scans of 20 ms have been used in this case) a fast trans-impedance amplifier is necessary. Under these conditions the photo-multipliers exhibit a characteristic noise process, they show large spikes with an essentially instantaneous onset and exponential decay. The amplitude of the spikes often exceeds the signal amplitude and the decay time depends on the decay constant of the trans-impedance amplifier. The cause of this type of noise is not known though it was found with all photo-multipliers tested.

The use of a look-ahead filter, followed by averaging of a small number of scans, followed lastly by a peak detection algorithm has been used here and shown to produce good results. Figure 6.5 shows an example where this technique has been applied to a relativity strong single mode signal to recover a much improved spectra.

The look-ahead filter works by detecting the fast onset of a spike and looking ahead three exponential decay constants to find the value after the spike. The signal inside the spike is then replaced by a linear interpolation between the values before and after.

The peak detection then works by performing a least-squares quadratic fit followed by a series of tests on the coefficients. The tests check whether each parabola is at a local maximum, the sign of the quadratic coefficient (indicates concavity) and if the peak is above a designated threshold. Note that this is the same procedure used in Section 3.5.1.

It is these identified peaks which are then stored and used in later calculations.

6.7. Data Acquisition Software

Data acquisition is primarily done with (but not limited to) National Instruments LabVIEW running on Red-Hat based GNU/Linux systems with National Instruments PCI data acquisition boards and NI-DAQmx Base drivers. LabVIEW provides a familiar rapid development environment for data acquisition and the GNU/Linux stack provides a high reliability environment with high resolution timing and clock synchronisation.

The general architecture of the standard data acquisition program follows the producer-consumer design pattern. A 'producer' loop obtains data at high speed and pushes it onto a queue. A second parallel 'consumer' loop pulls from the queue and processes the data. This de-coupling of the acquisition and processing ensures that acquisition will never be held up by periodic intensive processing. In most cases only the consumer loop will require modification as the needs of the experiments change.

The first stage of the processing is always a 'sanity check' which checks that basic properties of the raw waveforms (DC, RMS, frequency) fall within pre-specified ranges. Repeated failure of these sanity checks triggers a warning and causes the

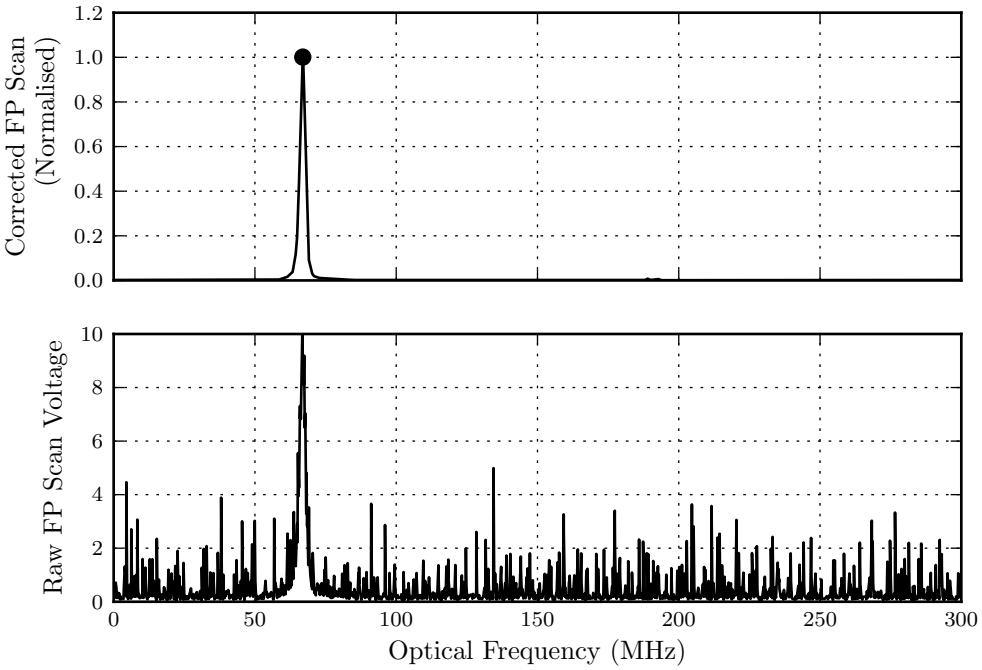


Figure 6.5. A Fabry-Pérot spectra showing single mode operation, showing the raw scan data (below) and a normalised scan following spike removal and averaging. The identified peak is plotted with a circle on the corrected spectra.

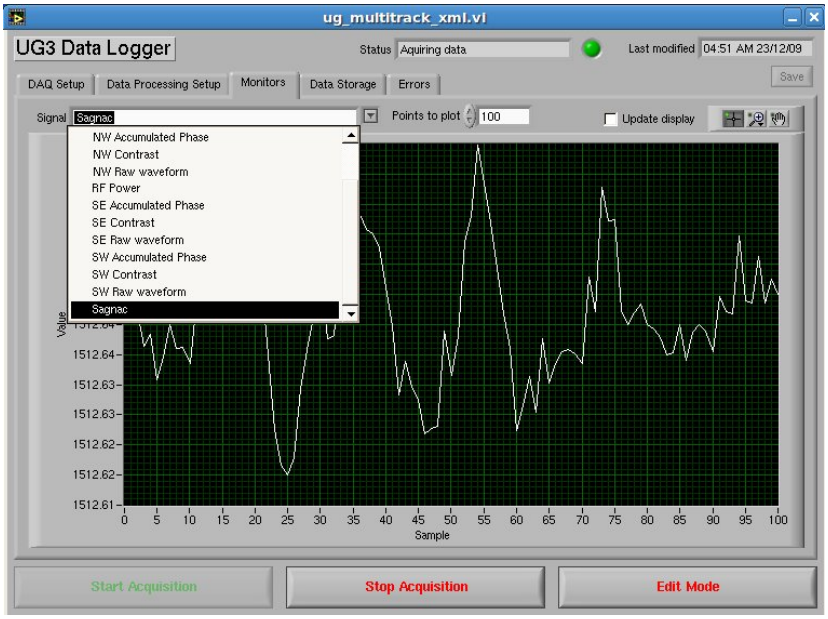


Figure 6.6. Typical data acquisition user interface.

processed data to be saved to a separate location for 'suspect' measurements and not automatically added to the database.

The standard user interface consists of a single window with tabs for configuring acquisition, processing and output. Another tab allows monitoring the data at various stages of acquisition, selectable via an automatically generated drop down box (Figure 6.6).

6.7.1. Communication

Ability to trigger acquisition of raw waveforms following the occurrence of a seismic event is provided by an XLMRPC interface implemented completely in LabVIEW. A seismometer system can call a 'new event' VI on another system via XMLRPC and pass the ID number of the event. This in turn sets a variable indicating the name of the event at the producer loop level. This variable is passed into the consumer loop via a shared queue, in the same way as the waveform data.

The consumer loop then behaves differently, saving all the raw waveforms and other data inside an `<event>` structure. It is often desirable to delay the consumption of waveform data slightly so that the data slightly preceding the seismic trigger will be

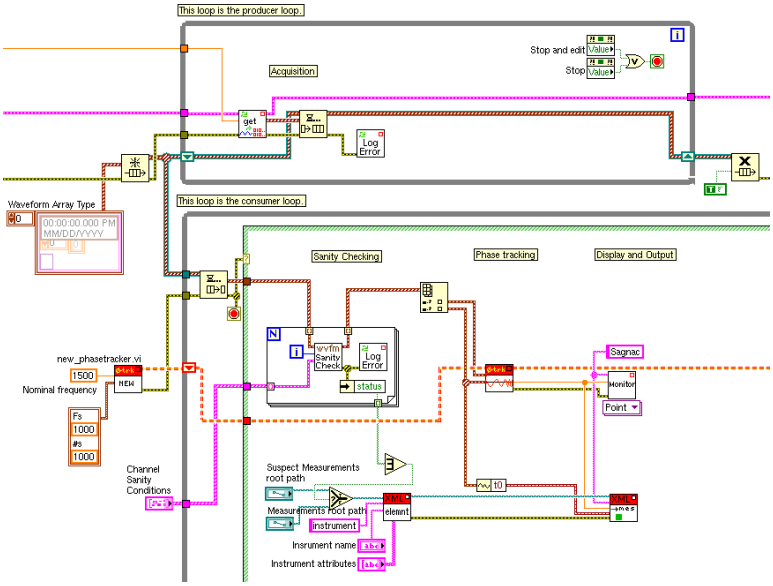


Figure 6.7. An example LabVIEW block diagram snippet showing the producer consumer architecture, phase tracking, XML output and monitoring.

saved. This functionality is provided by a delayed remove-from-queue VI which is a drop-in replacement for the standard LabVIEW remove-from-queue VI.

Once the data is merged to the database, a script (*reconcile_events*) will periodically search the GeoNet seismic catalogue to reconcile further event information create a `<eventinfo>` element for the event.

6.7.2. Constructing Output Files

While LabVIEW supports generation of XML files as a standard feature, this ability extends only to LabVIEW's own XML schema; generating arbitrary XML data is considerably more difficult. Third-party tools such as the EasyXML toolkit (JKI software) do exist, but were found to be cumbersome in our application.

As an alternative, a set of LabVIEW VIs were built for assembling the XML tree structure on disk using folders (for each child element) and files (containing element and attribute data). Shortcut VIs were made for assembling measurement elements specific to our application (see Figure 6.8). This approach minimises memory usage and also allows accumulation of raw data to be watched by looking at the tail of the file containing the elements for each sensor.

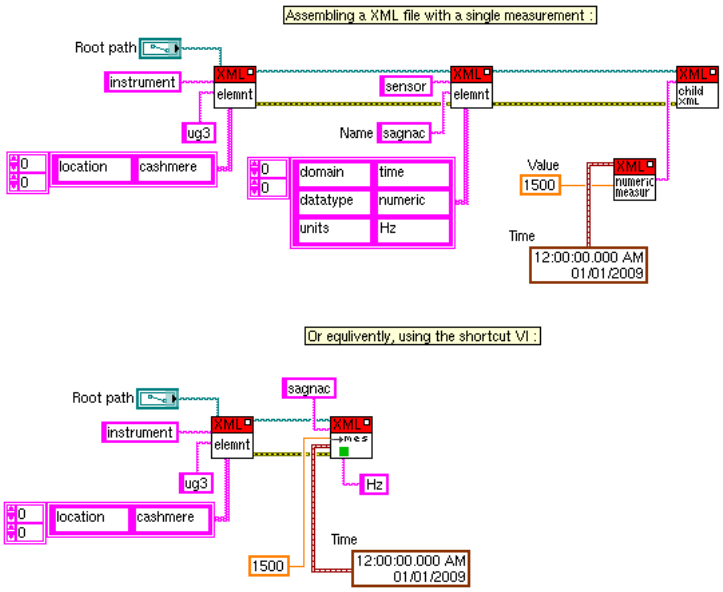


Figure 6.8. Assembling a ring laser XML file in LabVIEW.

Periodically the data acquisition system will run a script *buildxml* to flatten the directory structure to a single XML file[1]. This will then be check-summed and compressed (typically to around 3% of the original size) by bzip2 before being uploaded to the data storage system[2].

While the period in between generating data files is arbitrary, it is recommended to set this so that the individual compressed files are unlikely to grow much larger than about 5 Mbyte as scripts such as *buildxml* and *binaryseries* work by loading the whole document into a DOM object and there must be sufficient system memory to do this easily.

[1] As the XML assembly process can be relatively intensive, it is best to move the directory structure to a temporary location and run *buildxml* in a separate process (enclose in circular brackets on unix).

[2] The use of the `set -e` command is also recommended on Unix systems so that the building/uploading process will terminate if any error occurs.

6.7.3. Data Storage and Access

The data is stored in a repository running the eXist-db native XML database[1]. At the time of writing, eXist-db appears to be the most mature of the available native XML databases.

After data acquisition systems have uploaded compressed data files (with associated MD5 checksums) to a directory on the data storage system a number of housekeeping tasks are done before merging to the database. The data files must be checked against the supplied checksum, they must be run through the *binary_series* script which will convert any specified data to the `<binaryseries>` representation, and finally validated against the schema and merged to the database (*xmldbmerge* program). In the event of any problems, the problematic data must be set aside and the administrator contacted by email. All these operations are done by the *organise_incoming* script.

At this point we are able to analyse and extract data using queries written in the XQuery language. A simple time series can be obtained with just a few lines (the *universal_timeseries* script and associated web interface can be used to construct and execute such a query, and also has a number of useful features). Alternatively the expressiveness of XQuery can be fully taken advantage of to obtain results matching very specific criteria across multiple instruments and measurements.

An example of a rich query is one which looks inside autocovariance measurements (calculated hourly), finds the frequency interval corresponding to maximum amplitude and returns a time series of these maxima across multiple instruments. Figure 5.1 was created in this way and the corresponding XQuery at the heart of this analysis is just 9 lines.

Queries are typically processed in a few seconds to a couple of minutes, depending on the volume of data being returned. This level of performance is made possible by configuring eXist-db to pre-parse the numerical values and generate indexes which are then used by the query engine. Query time is mostly independent of the volume of data in the database.

[1] http://exist.sourceforge.net

```
1 let $col := collection("/db/rldata")/rldata
2 for $mes in $col/instrument[@name="ug3"]/sensor[@name="RF Power"]/measurement
3     let $time := xs:double( $mes/time )
4     let $value := xs:double( $mes/value )
5     where $time > 1268681668 and $time < 1269481952
6         order by $time
7         return
8             concat( xs:string($time) , "," , xs:string($value) , "&#10;" )
```

Listing 6.7 A simple XQuery example for producing a time series of RF Power measurements as CSV. Note the `
` string encodes a new-line character.

Instrument	Sensor	Type	Domain	Data type	Last recieved	Last value
ug3	NE Accumulated Phase		time	numeric	2 hours, 15 mins ago	-269088.5 cycles
ug3	NW Accumulated Phase		time	numeric	2 hours, 15 mins ago	-269509.98 cycles
ug3	RF Power		time	numeric	3 hours, 14 mins ago	2.71 V
ug3	NE Contrast		time	numeric	2 hours, 15 mins ago	0.58 ratio
ug3	SW Accumulated Phase		time	numeric	2 hours, 15 mins ago	-268998.24 cycles
ug3	SE Contrast		time	numeric	2 hours, 15 mins ago	0.4 ratio
ug3	NW Contrast		time	numeric	2 hours, 15 mins ago	0.69 ratio
ug3	SW Contrast		time	numeric	2 hours, 15 mins ago	0.54 ratio
ug3	Sagnac		time	numeric	2 hours, 15 mins ago	1512.5 Hz
ug3	SE Accumulated Phase		time	numeric	2 hours, 15 mins ago	-269423.72 cycles
ug3	autocovariance		frequency	numeric	167 days, 17 hours, 11 mins ago	0 rad/s/s
ug3	NE Raw waveform		time	numeric	19 hours, 8 mins ago	0.43 V
ug3	SE Raw waveform		time	numeric	67 days, 1 hour, 26 mins ago	4.12 V
ug3	NW Raw waveform		time	numeric	67 days, 1 hour, 26 mins ago	2.54 V
ug3	SW Raw waveform		time	numeric	67 days, 1 hour, 26 mins ago	3.12 V
ug3	2FSR-FS		time	numeric	93 days, 1 hour, 55 mins ago	7798466.37 Hz
ug3	2FSR+FS		time	numeric	93 days, 1 hour, 32 mins ago	7792408.45 Hz
g0	autocovariance		frequency	numeric	169 days, 8 hours, 11 mins ago	0 rad/s/s
g0	Top West Accumulated Phase		time	numeric	20 days, 5 hours, 25 mins ago	1373030.47 cycles
g0	Sagnac		time	numeric	20 days, 5 hours, 25 mins ago	288.87 Hz
g0	Top West Raw waveform		time	numeric	20 days, 6 hours, 23 mins ago	6.44 V
tiltmeter	EW Tilt		time	numeric	98 days, 10 hours, 40 mins ago	-0.47 urad
tiltmeter	NS Tilt		time	numeric	98 days, 10 hours, 40 mins ago	237.93 urad
EARSS	X Filtered		time	numeric	107 days, 20 hours, 26 mins ago	0 m/s
EARSS	Y Filtered		time	numeric	107 days, 20 hours, 26 mins ago	0 m/s
EARSS	Z Filtered		time	numeric	107 days, 20 hours, 26 mins ago	0 m/s
EARSS	X velocity		time	numeric	1 hour, 21 mins ago	0 m/s
EARSS	Y velocity		time	numeric	1 hour, 21 mins ago	0 m/s
EARSS	Z velocity		time	numeric	1 hour, 21 mins ago	0 m/s
c2	Sagnac Accumulated Phase		time	numeric	56 days, 2 hours, 17 mins ago	-19070.85 cycles
c2	Sagnac		time	numeric	56 days, 2 hours, 17 mins ago	79.49 Hz
c2	Sagnac Raw waveform		time	numeric	60 days, 1 hour, 46 mins ago	1.61 V

Figure 6.9. Web browser screenshot showing an example output from the instrument summary XQuery program.

```
<events>
{
let $alldata := collection("/db/rldata")/rldata
let $instrNames := distinct-values($alldata/instrument/@name)

(: Get all unique events :)

let $events := collection("/db/rldata")/rldata/instrument/sensor/event
let $eventNames := distinct-values($events/@name)
for $eventName in $eventNames
    let $thisEvent := $events[@name=$eventName]
    let $eventTime := min( $thisEvent/measurement/time/binaryseries/@min )

    (: Get the instruments in order of which this event arrived at first :)
    let $instruments := for $instrument in $alldata//event[@name=$eventName
        ]/../..
                        let $arrivalTimes := min($instrument//event[@name
                            =$eventName]//time/binaryseries/@min)
                        where $arrivalTimes > 0
                        order by $arrivalTimes
                            return $instrument

    (: Output the arrival time and details for the first instrument :)
    return <event>
                <arrivaltime> { $eventTime } </arrivaltime>
                <firstinstrument> { string($instruments[1]/@name) } </
                    firstinstrument>
                <lat>{ string($instruments[1]/@latitude)  }</lat>
                <lon>{ string($instruments[1]/@longitude) }</lon>
           </event>
}
</events>
```

Listing 6.8 A more involved XQuery example producing a custom output containing the details of the first instrument to record a seismic event. Note that the `(:...:)` format is used for comments in XQuery.

```
<events>
    <event>
        <arrivaltime>1.269470195792E9</arrivaltime>
        <firstinstrument>ug3</firstinstrument>
        <lat>-43.5746/>
        <lon>172.6232/>
    </event>
     ...
</events>
```

Listing 6.9 Output snippet of example Listing 6.8.

Ring Laser	GeoNet	Instrument	Date	Start time	Duration	Magnitude	Depth	Latitude	Longitude
eq1273740101	3306178	EARSS	2010-05-13	08:41:41	10 s	3.452	126.09 km	-41.35501	172.94904
eq1273740101	3306178	g0	2010-05-13	08:41:45	10 s	3.452	126.09 km	-41.35501	172.94904
eq1273740101	3306178	ug3	2010-05-13	08:41:39	10 s	3.452	126.09 km	-41.35501	172.94904
eq1270963922	3289433	EARSS	2010-04-11	05:32:02	26 s	2.51	167.1719 km	-41.44941	172.38188
eq1270963922	3289433	ug3	2010-04-11	05:32:00	5 s	2.51	167.1719 km	-41.44941	172.38188
eq1270960889	3289418	EARSS	2010-04-11	04:41:28	21 s	2.475	45.9129 km	-40.04544	176.60376
eq1270960889	3289418	ug3	2010-04-11	04:41:27	4 s	2.475	45.9129 km	-40.04544	176.60376
eq1270426764	3286275	EARSS	2010-04-05	00:19:24	10 s	2.302	105.912 km	-37.80278	176.53606
eq1270426764	3286275	ug3	2010-04-05	00:19:22	2 s	2.302	105.912 km	-37.80278	176.53606
eq1270426019	3286015	EARSS	2010-04-05	00:06:58	27 s	2.927	23.9834 km	-39.65233	176.81721
eq1270425993	3286015	EARSS	2010-04-05	00:06:32	10 s	2.927	23.9834 km	-39.65233	176.81721
eq1270321828	3285425	EARSS	2010-04-03	19:10:27	44 s	3.825	33 km	-43.24669	172.7632
eq1270321828	3285425	ug3	2010-04-03	19:10:27	7 s	3.825	33 km	-43.24669	172.7632
eq1269446302	3280290	EARSS	2010-03-24	15:58:21	12 s	3.08	65.0173 km	-42.32197	172.96027
eq1269446302	3280290	ug3	2010-03-24	15:58:20	2 s	3.08	65.0173 km	-42.32197	172.96027
eq1269436198	3280221	EARSS	2010-03-24	13:09:58	9 s	2.646	29.3493 km	-40.83144	175.74294
eq1269436198	3280221	ug3	2010-03-24	13:09:57	2 s	2.646	29.3493 km	-40.83144	175.74294
eq1264185096	3246638	g0	2010-01-24	15:26:45	1 s	2.462	5 km	-38.77805	179.41115

Figure 6.10. Web browser screenshot showing an example output from the (reconciled) event information XQuery program. The GeoNet ID provides a convenient link to the QuakeML resource on the GeoNet website, while the instrument link directs to the web interface of the universal_timeseries script with the timing and instrument information already filled in for the respective instrument.

6.8. Summary

The key advantages of using a standardised XML data format are that we need only write the routines to generate and process the data once and they can then be used in many different contexts. Additionally, having a machine readable specification allows for automated validation which in turn allows for a good level of assurance that the data gathered is not ambiguous. This allows problems to be discovered early and should result in greater volumes of higher quality data being gathered.

A key advantages of using a native XML database for storage is that we can write queries to obtain data based on the mathematical relationships between data sets and discover relationships between data without knowing *a priori* when the data was gathered, or on what instrument. Unlike with a traditional relational database, we store the data in the same format it was acquired in. This means that users only have to understand one data structure.

7. Conclusion and Further Work

In the earlier sections of this thesis we summarised a number of fundamental and practical problems which limit the ultimate performance of ring laser gyros. We have also summarised a number of potential scientific applications to which a ring laser with ultimate performance could make an important contribution.

The goal of this chapter is to summarise the key results and ideas which have through this work been shown to have potential for reducing limits on performance as well as those which were unsuccessful. These individual conclusions however are not useful in the context of a wider view on the potential of a next generation ring laser gyro. To this end we combine what has been learned with general operational experience and an assessment of the practicality and scientific merit of potential measurements. This allows us to make a set of recommendations for a next generation ring laser gyro. It also allows us to take an analytical approach to determining good potential sites for such an instrument.

7.1. The Limitations of Laser Gyros

The physical mechanisms which are well known to limit the performance of laser gyros are:

1. Variations in area.
2. Variations in perimeter.
3. Variations in orientation of the laser and site relative to the axis of rotation being measured.
4. Variations in optical frequency caused by backscatter, which ultimately can cause phase locking in smaller devices.
5. Unbalanced flows in the plasma.

6. Variations in dispersion of the plasma.
7. The quantum limit to rotation rate measurement.
8. Size limitations imposed by limitations in super-mirror technology and diffraction.

Variations in the area of laser gyros can be eliminated to some extent by building a monolithic device, although this approach does not scale well to very large lasers. Correcting changes in the relative side lengths of laser gyros to interferometric accuracy has been successfully demonstrated in this work through the technique of *multiple phase tracking* (Section 4.2) provides a small correction to the Sagnac frequency which could well be usefully applied to a next generation ring laser gyro.

Multiple phase tracking does not fully allow area changes of a rectangular laser to be corrected for because it is not sensitive to diagonal shear deformations (Section 4.1), for which we need to measure the angles of incidence. We note that this limitation is eliminated in a gyro built with a triangular configuration (Section 4.1.1). Measuring angles of incidence has been experimentally demonstrated at the proof-of-principle level by the method of *contrast ratio tracking* (Section 4.3). We note that a fundamental limitation of this technique is the inability to determine the direction of angle change relative to the nominal angle.

Correction for variations in the orientation of the plane of the laser to the axis of rotation has been demonstrated previously. This was done by directly measuring the tilt of the laser plane [19]. Ideally however, this problem could be reduced to a 2nd-order effect by building the laser gyro with the plane perpendicular to the rotation axis.

Variations in perimeter can be partially eliminated by building a monolithic device. However this is not practical for ultra-large devices. The perimeter can be found from measurements of the mode-spacing (FSR). This can be exploited by operating laser gyros using the *alternate-split* technique (Section 4.5). This technique has been experimentally demonstrated with perimeter measurements to sub-wavelength precision. Switching times of the order of 5 minutes have been demonstrated. This effectively allows simultaneous measurement of rotation and perimeter when considering effects varying on the time scale of hours or longer, for example tidal strains. The main problem encountered in this proof of concept experiment was the accuracy

to which the overall Sagnac frequency could be determined. However this is mostly an instrumentation problem and we expect progress could be made following further development of this technique.

Frequency pulling due to backscatter can be eliminated by operating laser gyros using the *alternate-split* technique. This has been experimentally demonstrated at the proof-of-principle level, though more accurate frequency measurement is required for rotation rate measurements made in this way to be competitive with the regular single-mode Sagnac measurements. Such improvements may be afforded by lower noise photo-detectors and higher speed data acquisition electronics.

Null shifts caused by both unbalanced plasma flows and scale factor shifts caused by dispersion in the plasma can be eliminated by *in-ring-down measurement* (Section 4.7). The experimental technique and a robust analysis technique has been demonstrated. This mode of operation has not yet been demonstrated to be competitive with regular continuous-wave Sagnac measurement due to noise in the photo-detectors and the relatively short (particularly compared to what might be possible with a new generation of mirrors) ring-down time of the UG-3 cavity.

In-ring-down measurement is best applied following *mode seeding*. This has been demonstrated at the proof-of-principle level (Section 4.6). It has been shown experimentally that a seed beam with intensity sufficient to excite just a few resonant photons in the cavity is sufficient to successfully induce seeding, provided that the seed beam is sufficiently well matched in frequency and alignment.

The maximum size of possible laser gyros is likely to continue to increase as super-mirror fabrication technology continues to improve. It is also likely that further increases in maximum optimal scale factor can be obtained by developing a system to automatically maintain alignment such that the laser beams are always positioned in the clear aperture of the laser mirrors.

Another possibility for making progress on the size limitation is decreasing the wavelength of the laser (Section 2.4.5). Operation on the 543.3 nm transition of neon effectively increases the scale factor (directly) by 14% and increases the maximum optimal scale factor by 20.6%. The practical aspects of operation at this wavelength present a minimal departure from existing experience, though other less familiar systems operating in the UV would give a more dramatic increase in scale factor and

thus are interesting options.

Indirect limitations in the performance of laser gyros arise through deficiencies in understanding aspect of the gain and excitation process. Some progress has been made to this end in Chapter 3, although there is still much to be learned. In particular, the reliance on empirical models for the saturation and broadening processes needed to calculate the gain curve is less than ideal. The lack of a model for predicting the saturation intensity to good accuracy is a particular deficiency at present.

These limitations in understanding affect performance indirectly. For example if the gain processes could be sufficiently well understood so that an 'ideal' gas mixture could be found, then the gas lifetime could be improved and consequently various temperature dependent null shifts could be minimised. In addition, an accurate gain curve is necessary for calculations of the frequency pulling due to dispersion.

7.2. Integration of New Techniques

In the previous section we summarised techniques which, when taken individually offer hope for eliminating all of the factors presently known to limit the performance of laser gyros.

However, these techniques can not necessarily be applied simultaneously and further work is needed to move them beyond the proof-of-principle level.

In particular, multiple corner phase tracking and FSR tracking can not be used simultaneously as the latter requires split-mode operation. The alternate-split technique allows both perimeter and rotation measurement, though no area measurement. Measurement of relative side lengths (and in turn area if operating a triangular configuration) could be provided by one or more auxiliary lasers (see Figure 7.1) though a laser of this design has not been demonstrated before.

To fully account for changes in geometry when using the alternate-split mode technique, switching must occur at a frequency faster than any mechanisms which perturb the geometry of the laser. If earthquakes are ignored, this means a switching period of around 200 ms is necessary. A control technique based on measurement of spectral centroids has been demonstrated to achieve switching periods of around 300 s (Section 4.5).

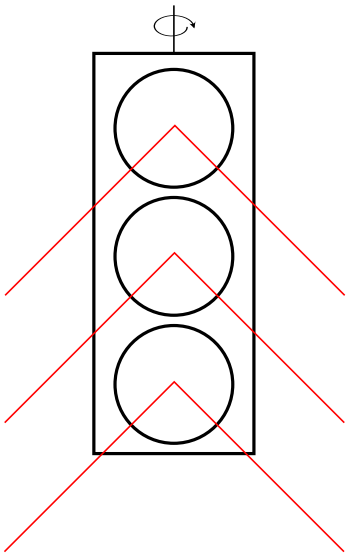


Figure 7.1. Possible implementation of a mirror holder in a ring laser with two auxiliary reference beams.

	Area	Perimeter	Sagnac (backscatter effected)	Sagnac (ideal)
Single-mode	✓	✗	✓	✗
Split-mode	✗	✓	✗	✗
Alternate-split	✗	✓	✗	✓

Table 7.1. Comparison of the capabilities of proposed new techniques for ring laser operation.

The technique of *injection seeding* (Section 4.6) is the most promising for fast control of the splitting configuration. What remains to be experimentally demonstrated is a technique for maintaining seed beams in each of the desired splitting configurations and successfully injecting the beams into a large ring cavity. An experiment to demonstrate this has been outlined in Section 4.6.2.

If a fast seeding technique is used, the next logical step is to apply in-ring-down measurement to the FSR$\pm f_s$ signals. As these frequencies are much higher than the Sagnac frequency, accurate measurement of frequency is easier as there are many more cycles in one ring-down. However the fraction of the measured frequency corre-

sponding to the Sagnac frequency is small, so the requirements for accurate frequency measurement are overall still very high. Further investigation into low-noise photo detectors and high speed data acquisition electronics is necessary to achieve good results with this technique.

If fast injection seeding with in-ring-down measurement can be achieved, the next logical step is to reduce the relative amplitude of quantum phase fluctuations by increasing the total photon number. The simplest way of achieving this is to increase the single-mode power by choosing a high gas pressure to increase the multi-mode threshold and correspondingly increasing the length of the gain section.

We note that a preliminary investigation of the upper limit to single mode power is possible with the installation of an extended gain section in existing equipment (UG-3). This would also allow investigating operation on the lower gain 543.3 nm transition.

To properly model the gain curves resulting from different gas mixtures a good knowledge of how the saturation intensity in the helium-neon system behaves under different conditions is required. Some progress on this has been made as a phenomenological model of the saturation intensity as a function of gas pressure and mixture has been developed based on a large number of measurements of the multi-mode threshold (Section 3.7). However this does not take into account variations in gas temperature, which constitute a previously unrecognised but important parameter. Repeatable measurements of gas temperature have proven difficult.

A related area much in need of further work is the generation of the plasma. This is presently achieved with RF excitation via a passive impedance matching circuit. Designing these matching circuits is a difficult task; a trade-off must always be made between the ability to provide sufficient field strength to start the discharge and to deliver power with sufficient efficiency to the running plasma. Inevitably, a significant amount of the RF power is radiated, and this RF interference is particularly troublesome when trying to make precise measurements. RF interference is a particular problem for a ring laser in the application of alignment control for a gravity wave detection interferometer; these facilities are very strict about minimising sources of interference.

Further work on the excitation process is also necessary if operation using mode-

seeding and in-ring-down measurement are to be further explored. The delays necessary to maintain a stable plasma are an important factor in techniques which require the plasma to be restarted regularly. Additionally, these techniques place very high demands on detector noise. Lastly, it is likely that further investigation into the excitation process will lead to insight into techniques to minimise asymmetries in the discharge. Such asymmetries may lead to unbalanced plasma flows and Sagnac frequency pulling.

One additional area of further work which is necessary before true wavelength-level perimeter measurement is possible is to understand where reflection can be considered to occur on the multi-layer dielectric mirror coatings. This is something not generally well known, since in other applications of these types of mirrors (i.e. gravity wave detection) it is only the relative phase between the incident and reflected beams that is of importance; the absolute positions of the optical components are not known to sub-wavelength precision [142].

7.3. Further Measurements

There are a number of geophysical measurements for which UG-3 and associated instruments are uniquely suited.

While originally intended as correction signals for rotation measurement, the relative side length measurements of UG-3 are useful for measuring earth strain. This can then be used to calibrate the many parameters used in calculating tidal strains. These parameters are likely to become increasingly interesting to the geophysical community, following the recent article by Thomas *et al.* [94] in which a correlation was drawn between earthquake probability and tidal strains.

The existence of tidal rotations is an interesting prediction stemming from the modelling of tidal strains and tilts. The tidal rotation signal predicted for UG-3 is small, though potentially measurable with a sufficiently long data set.

The combination of orthogonal laser gyros UG-3 / G-0 and a sensitive seismometer in the same location provide a good opportunity to study the geophysical origins of microseismic signals. Figure 5.1 for example shows how a correlation can be identified between microseismic signals in instruments designed to measure rotation about two orthogonal axes. This suggests a degree of cross coupling between the two degrees of

freedom. Determination of the physical magnitude of the coupling could give insight into the origin of the effect.

Analysis of geophysical phenomena such as those given above has been made much easier following recent standardisation of data acquisition and storage systems. A set of reusable components has developed which allow data acquisition systems to be implemented quickly and readily modified as the needs of the experiments change. This was necessary to improve the maintainability and efficiency of the overall data life-cycle from acquisition through to publication. Important features of the system now implemented include a high level quality control, ability to do rich queries across all data to discover relationships with minimal effort, and the ability to easily transform data into alternate representations. It is hoped that this improved accessibility will allow for improved collaboration with a wider research community.

7.4. A Next Generation Ultra-Large Ring

In Chapter 5 we detailed several potentially interesting measurements that might be made with a next generation ultra large ring laser gyro.

If designed with a Lense-Thirring experiment in mind, such an instrument appears to have a plausible chance at successfully making this measurement. This would constitute an important confirmation of general relativity. Because of the scientific importance of the measurement, and the relatively large signal (when compared to other potential applications) this is probably the most promising potential application of a next generation ring.

Another application which stands out is the measurement of coseismic rotations. In this application, quite large improvements in the probability of observing an event can be obtained with relatively modest improvement in performance (see Figure 5.3).

7.4.1. Features

Given what we have learnt regarding the operation of ring lasers we are in a good position to determine the important features of a next-generation ultra large ring.

1. The laser should be symmetrical and have *perimeter of the order 100 m*.
2. The laser should be built with the plane *perpendicular to the earth rotation*

axis. This will eliminate to first order the effect whereby changes in tilt of the laser plane change the observed rotation signal. It will also maximise the rotation rate being measured.

3. The laser should use an *equilateral triangular geometry*. This will eliminate the insensitivity to some components of shear and allow both area and perimeter to be measured to interferometric accuracy. A triangular geometry ensures the laser plane is always well defined which in turn ensures the polarisation of the output beams is well defined. In addition, one less mirror reduces the backscatter and associated Adler pulling and also reduces total losses and instrument cost when compared to a rectangular configuration.

4. The laser should be built with reasonable laboratory space at at least two of the corners. If detector noise can be made sufficiently small, this will allow the equipment necessary to operate the laser using the *alternate-split* technique with fast *injection seeding*.

5. The laser should employ *secondary reference beams* as a means of establishing relative side lengths and enabling continuous Sagnac measurement in the case that the alternate-split technique is used.

6. Emphasis should be placed on obtaining a cavity with the highest possible Q, and ideally ring-down of the order 10 ms. This could allow measurements to be usefully made using the *in-ring-down measurement* technique. This will require an improvement in mirror technology; likely significantly more attention to the super-polished aperture and the deposition process for the Bragg stack will be required.

7. A feedback system[1] should be developed to keep the laser beams in *exactly the clear aperture* of the laser mirrors. This will require the laser mirrors and combining optics being rigidly mounted together on a optical breadboard. This will eliminate the ambiguity with present lasers whereby the combining optics and beam position sensors can move independently of the beam.

8. A *mode control aperture* should be installed in the cavity to allow discrimination of high order transverse modes. This becomes increasingly important as super-

[1] Possibly made using quadrant photodiode detectors and piezoelectric actuators.

mirror technology continues to improve. Cavities with higher Q will allow higher order transverse mode configurations to operate that would previously have been below threshold.

9. Large diameter *isolating gate valves* should be installed on either side of the laser mirror enclosures, gain tube and any other important sections. It should be possible to evacuate each isolated section individually. This will allow the cavity to be maintained free of contaminants because it will be possible to change components without raising the whole cavity to atmospheric pressure. This will also significantly improve operational flexibility because vacuum problems will be easier to troubleshoot and (if necessary) it will be possible to bake individual sections to remove contaminants.

7.4.2. Location

Perhaps the most important design consideration for a ring laser incorporating the above design features is the location. The location should be selected taking into consideration the magnitude of the frame-dragging effect, the stability of the instrument and cost.

Given this, the analysis of potential sites to determine suitability is an interesting exercise. The overall suitability of a site can be determined by an appropriate weighting of a number of metrics which can be calculated from readily available databases.

The geographical suitability of potential sites is first determined by obtaining the elevation and eliminating sites with elevation below 5 m (likely to be under water). The elevation E of remaining sites can be used to determine a metric which eliminates mountainous regions

$$m_{\mathrm{E}} = 1 - \frac{E_{\phi,\theta}}{\max(E).} \tag{7.1}$$

Elevation data was obtained from the GEODAS system[1]. For each point in question the elevation was found from the nearest point on the 10×10 arcmin survey.

The magnitude of the frame-dragging effect can be calculated from Equation 5.14. Here we are only interested in the relative magnitude as a function of latitude ϕ. We assume the laser is built perpendicular to the earth rotation axis, ignore the sign of

[1] Maintained by the National Geophysical Data Center, Boulder, Colorado, USA.

the effect and normalise it from 0...1. We define the metric

$$m_{\mathrm{fd}} = \left| \frac{3\sin^2\phi - 1}{2} \right|. \tag{7.2}$$

The stability of the instrument can be partly determined by the seismic activity of the region. This was found from the global seismic hazard map[1]. This is produced by a collaboration of hundreds of scientists from around of the world and gives a value h from 0...10 based on the peak ground acceleration that has a 10% chance of being exceeded in the next 50 years [143]. We use this to define the metric

$$m_{\mathrm{s}} = 1 - \frac{h}{10}. \tag{7.3}$$

An additional factor affecting the stability of the instrument is the variability of local weather. Weather systems causing variations in atmospheric pressure have been a major concern in all lasers built to date. While active pressure control has been demonstrated (C-II, G) this is likely to be very expensive for an ultra-large ring.

A measure of the stability of the atmosphere has been obtained by finding the quadratic mean (RMS) of monthly temperature anomalies[2] from 1850 to 2009 at the location in question (denoted $R_{\theta,\phi}$). This data has been extracted from the CRUTEM3 [144][3] data set. Points missing from this data set are obtained by linear interpolation from neighbouring points. Suitability of the weather at a site is determined by the metric denoted

$$m_{\mathrm{w}} = 1 - \frac{R_{\phi,\theta} - \min(R)}{\max\left(R - \min(R)\right)}. \tag{7.4}$$

A simple way to estimate the relative cost of an instrument at a given site is to estimate how remote the site is. Populated areas are likely to have existing transportation and construction infrastructure which would reduce the cost of construction and logistics. Most of the world population is concentrated in a small number of regions of very high density and the distribution shows a strong exponential tendency.

[1] Published by the U.S. Geological Survey, Swiss Seismological Service, International Lithosphere Program and available online from http://www.seismo.ethz.ch/GSHAP/index.html.

[2] Monthly anomalies are defined as the maximum departure from the 1961–1990 mean recorded in a given month.

[3] Provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA.

Metric	A	B	C
m_{fd}	0.5	0.3	0.0
m_{e}	0.05	0.05	0.1
m_{s}	0.25	0.15	0.35
m_{w}	0.1	0.25	0.35
m_{p}	0.1	0.25	0.2

Table 7.2. Different sets of weighting parameters used in the calculations shown in Figures 7.2, 7.3, 7.4, 7.5, 7.8, 7.9.

Highly populated areas offer no advantage for construction of a ring laser, so population density data p (obtained from the GPWV3, 1995 data set[1]) was transformed to a linear distribution and normalised, giving the metric

$$m_{\mathrm{p}} = \frac{\log(p+1)}{\max(\log(p+1))}. \tag{7.5}$$

While it is true in general that populated areas are likely to have elevated seismic noise due to human activity, this is likely to be quite localised. Here we are more concerned with the broader picture and assume that fine-tuning of the site could minimise noise caused by human activity.

These metrics have been calculated for 20,000 points evenly distributed across the surface of the earth following the distribution method of Saff and Kuijlaars [145]. The metrics have then been weighted by factors given in Table 7.2. Results for three different weighting sets have been calculated. The resulting overall suitability metric has in each case been colour-mapped and overlaid on a projection of the world coastlines and geopolitical borders. Both equal-area Mollweide projections (Figures 7.2, 7.4, 7.8) and northern and southern polar Lambert azimuthal equal-area projections (Figures 7.3, 7.5, 7.9) are shown. Figures 7.6 and 7.7 are enlargements of Figure 7.4 for southern Asia and norther Europe (azimuthal equidistant projection). In all figures, dark red represents the most desirable locations and blue the least desirable.

Let us first consider an ultra large ring laser gyro intended for measuring frame

[1]Maintained by the Center for International Earth Science Information Network (CIESIN), Columbia University; and Centro International de Agricultura Tropical (CIAT).

dragging, and for which cost is of little concern. This is represented by weighting set A. We see that only the most inaccessible polar regions show up as particularly desirable (∼0.8). We note that a device constructed in these locations is likely to be very expensive, and it may be the case that multiple instruments in slightly less desirable locations could be constructed at similar cost. Let us now consider the reasonably desirable (∼0.7) locations. Northern Europe, in particular western Iceland, most of Norway, western Sweden and northern Scotland do stand out. The latter is notable for high desirability given the relatively large distance from the pole. However we note that the advantage these locations present over many slightly less desirable locations (0.55 to 0.65) is relatively slight.

Let us now compromise slightly on the magnitude of the frame dragging effect, instead favouring more atmospherically stable and more accessible areas. This is represented by weighting set B. As the relative weightings are changed towards these values, we note the above mentioned northern European regions become less desirable while the more central European locations increase in desirability significantly compared to the rest of the world. Simultaneously, a large number of equatorial regions show up with similar or higher desirability. Several regions scattered across central Africa show up, however due to political instability these can be eliminated. This leaves several specific locations across south-east Asia and southern India.

Overall we see that an ultra large ring laser gyro constructed near the equator is at least as compelling as a polar device, unless ultimate performance is to be obtained at any cost. This is because the reduction in magnitude of the frame dragging effect is approximately cancelled by other factors leading to increased stability. However the desirability of equatorial locations can vary significantly over small distances, so optimising the location is important.

Finally, let us consider the case in which measurement of frame dragging is not considered a priority. This is represented by weighting set C. Here the values are chosen to give a high stability instrument, ideal for long term geodetic measurements. The main effect of this is to elevate regions in the mid-latitudes that were formally undesirable to becoming moderately desirable. However, the overall pattern of optimal locations remains surprisingly similar. This indicates that choosing a location ideal for measuring frame-dragging is not likely to be a significant impediment to other operations in the event that measurement of frame dragging is eventually shown to

be unachievable.

Overall, we note that while the specific weightings shown in Table 7.2 and some of the parameters used in calculating the individual metrics are quite arbitrary, the general pattern is fairly insensitive to the exact values of these parameters. In each case there are a few locations which consistently show up as very suitable for constructing a next generation ring laser gyro.

7.5. Final remarks

Over the past 15 years, ring laser gyro research has continued on a steady path building larger devices. The goal has been continual development and increase in performance. As we have seen from experience with UG-2, it is no longer sustainable to simply continue building larger and larger devices. It is now necessary to design ring laser gyros with the specific application and concrete program goals in mind from the outset. These determine the best operational technique to use, the optimal size and location.

In this thesis we have outlined a number of interesting scientific goals which ring laser gyros have a very real potential to make important contributions to if the presently understood limitations to performance can be minimised. Several new methods for operating ring laser gyros to achieve this have been proposed and experimentally demonstrated at the proof of principle level.

Despite its limitations, the UG-3 laser can be operated for long periods and produces useful data on a number of interesting geophysical effects. Most of these effects occur over long periods (e.g. tidal effects, polar wobble) and a very useful data set could be obtained from continuous observations with the UG-3 laser over a period of a year or more.

Such long-period observations have in general not been demonstrated with this instrument. Many corrections that could be applied to the data (for example those based on side length measurement) require continuous integration. Interruptions, particularly those longer than the 12 hour tidal cycle are highly disruptive because it is not possible to interpolate between the missing data. If long periods of 'observatory mode' data are to be obtained, it will be necessary to make a clear commitment to quickly resolve the technical problems which inevitably arise.

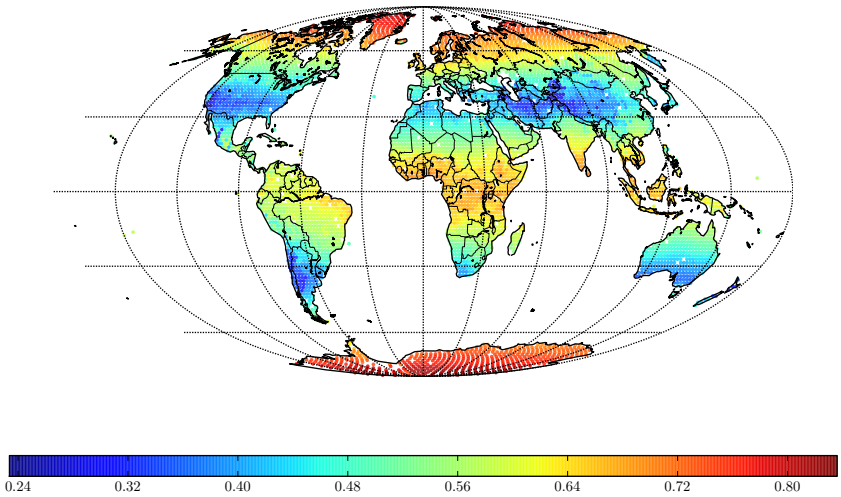


Figure 7.2. Weighting set A, Mollweide projection.

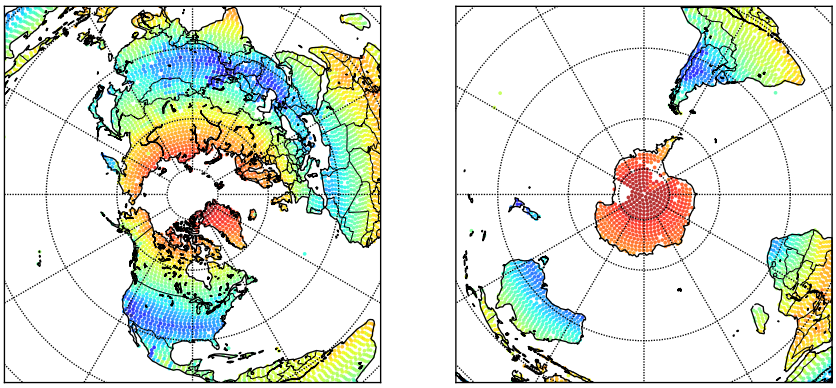


Figure 7.3. Weighting set A, polar Lambert azimuthal projections. Colour map follows Figure 7.2.

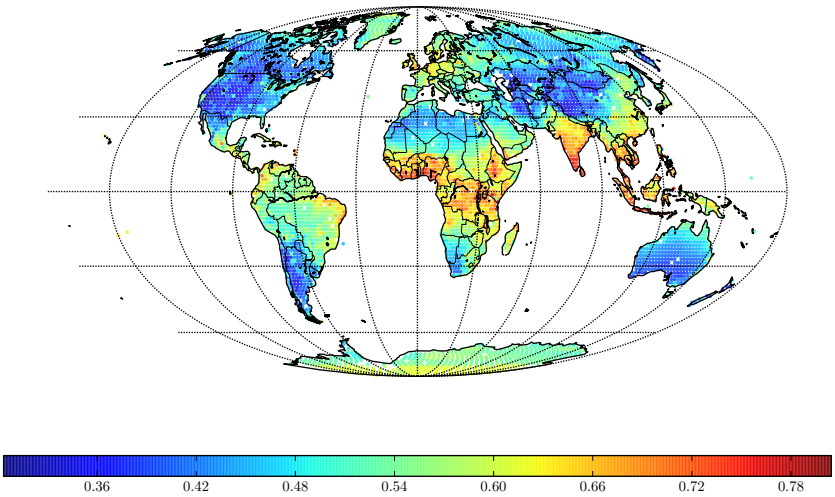


Figure 7.4. Weighting set B, Mollweide projection.

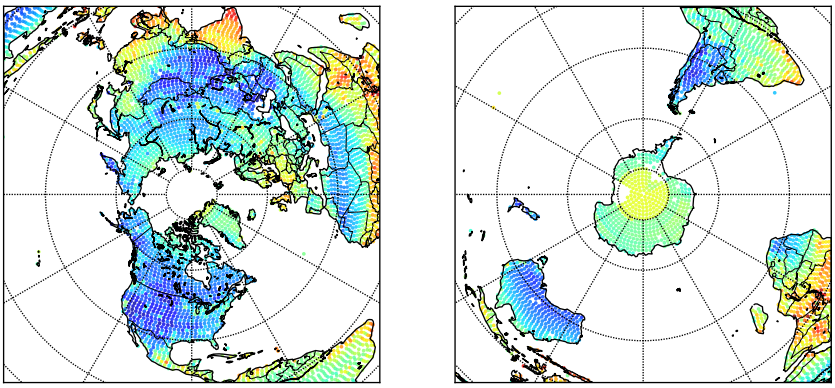


Figure 7.5. Weighting set B, polar Lambert azimuthal projections. Colour map follows Figure 7.4.

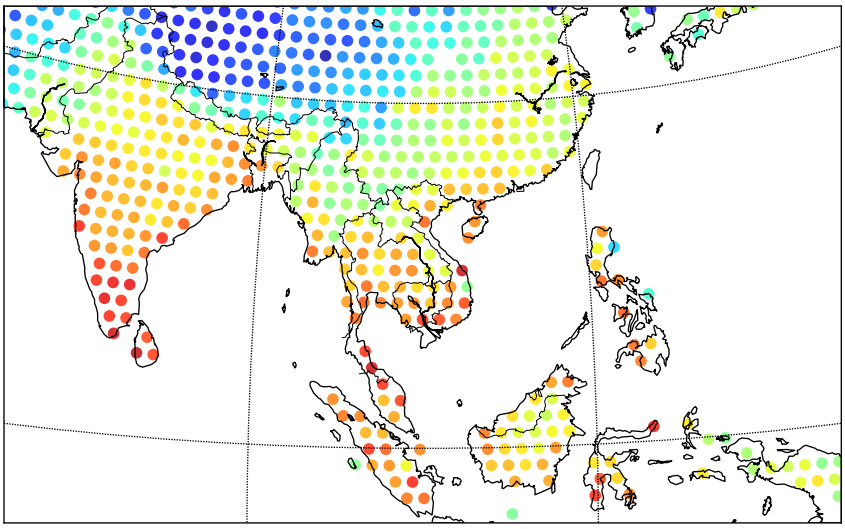


Figure 7.6. Weighting set B, azimuthal equidistant projection for south Asia.

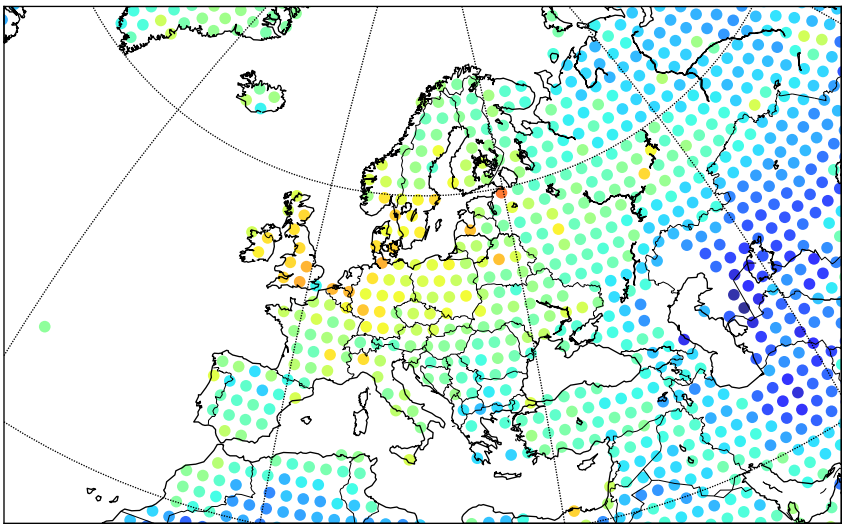


Figure 7.7. Weighting set B, azimuthal equidistant projection for north Europe.

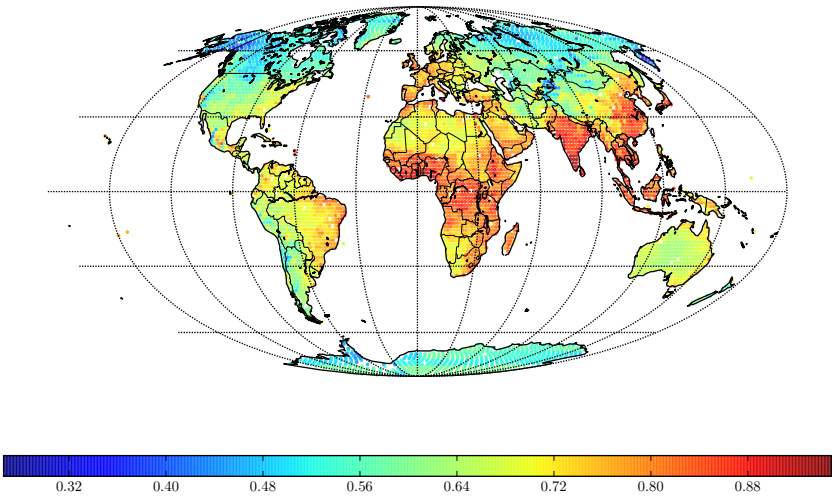


Figure 7.8. Weighting set C, Mollweide projection.

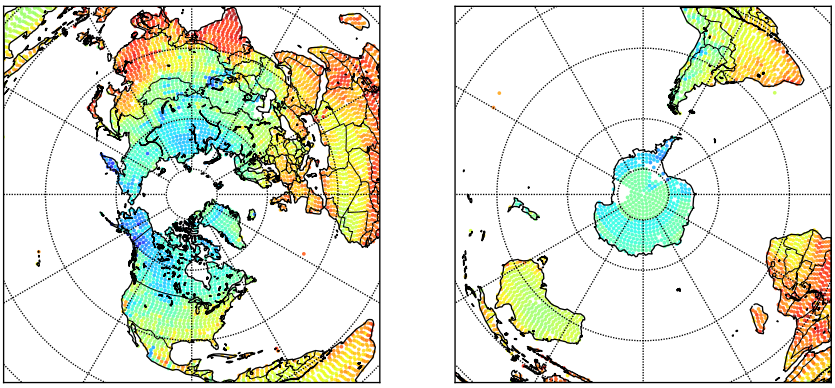


Figure 7.9. Weighting set C, polar Lambert azimuthal projections. Colour map follows Figure 7.8.

A. Data Analysis Software

A.1. Gaincurve module

A.1.1. Synopsis

The purpose of this module is to simplify the calculation of gain curves. It was written to incorporates the equations and various numerical models presented in Chapter 3. It makes use of the *physconst* module.

The module contains a base class/interface *gaincurve*. This defines the specification for any object which defines the gain at a particular optical frequency according to some phenomina or operation. The most important method it contains is the *gain(f)* method which takes a frequency in MHz and returns some measure of the gain. This method is aliased to the objects *__call__()* method, which allows gaincurves to be called as functions. This is particularly useful because it allows the user to pass a gaincurve object directly to standard optimisation routines. Additionally, all *gain(f)* methods can be passed either floating point values or arrays of floats and the and the returned type will adapt automatically. This is useful in a wide variety of contexts. The base gaincurve class also contains some useful general purpose methods, such as maximisation and evaluation over a frequency range.

A *gaincurve* object would not normally be instantiated directly; the *gain(f)* method does nothing and is intended to be overwritten by methods in inhereted classes. Several inherited classes are available, representing *Gaussian*, *Lorentzian*, *Saturation* and *Voigt* functions. How these are instantiated varies depending on the class. Most will require an instance of a *LaserGas* class which manages a gas mixture and automatically sets the various instance variables according to the gas mixture (note that these can always be overridden later). A *Voigt* gaincurve requires the component *Gaussian* and *Lorentzian* gaincurves when it is constructed.

Additional classes *Sumcurve* and *Convolve* are available. These take lists of other gaincurve objects when initialised, and when their *gain(f)* methods are called, they access the respective methods of the curves they were originally given and apply a mathematical operation to them, (i.e. summation, multiplication). A *Scalecurve* is also available which simply scales a given curve by a given floating point value.

All *gaincurve* objects are compatible with both continuous and discrete evaluation. Sometimes additional information is required to accommodate this, for example an optional parameter *discrete* is provided when constructing a *Voigt* gaincurve. When given an integer k, the *gain(f)* method will adapt to evaluating the integral discretely using Romberg integration on $2^k + 1$ pre-sampled evenly spaced points. This can be an order of magnitude faster, but introduces descretisation artefacts, as illustrated in Figure 3.4.

In many cases information about the gas mixture will be needed. A *LaserGas* class is provided which manages a gas mixture, allowing isotope species to be added at various pressures and the lasing transition to be specified. The various parameters are then found automatically using the *physconst* module.

Overall, the approach implemented here is useful because it allows gain curves to be built up from elementary functions each of which can be checked individually at any stage. The final gaincurve is evaluated through a call-chain to the various *gain(f)* methods. Because the resulting function is (optionally) continuous, it can be passed to optimisation routines for example to solve I/I_{sat} at the multi-mode threshold without being limited by numerical precision. Additionally, because it only needs to be evaluated for individual frequencies, performance when simulating the evolution of individual modes is greatly increased compared to earlier implementations by the author which relied on discretisation for all calculations.

A.1.2. Functions

savitzky_golay(*y*, *window_size*, *order*, *deriv*=0)

This is a utility function for smoothing (and optionally differentiating) data using a Savitzky-Golay filter.

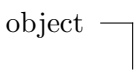
The Savitzky-Golay filter removes high frequency noise from data. It is well suited to smoothing noisy data and has the advantage of preserving the

original shape and features of the signal better than other types of filtering approaches, such as moving average techniques.

The main idea behind this approach is to make a least-squares polynomial fit of high order over an odd-sized window centred at each data point.

`y:`	the values of the time history of the signal. (numpy array or similar)
`window_size:`	the length of the window. Must be an odd integer number. (int)
`order:`	the order of the polynomial used in the filtering. Must be less then 'window_size' - 1. (int)
`deriv:`	the order of the derivative to compute (default value of 0 means only smoothing) (int)

A.1.3. Class Gaincurve



gaincurve.Gaincurve

Available sub-classes: gaincurve.Convolve, gaincurve.Derivativecurve, gaincurve.Gaussian, gaincurve.Lorentzian, gaincurve.Saturation, gaincurve.Scalecurve, gaincurve.Sumcurve, gaincurve.Voigt

This is the base class representing a gaincurve. Normally a gaincurve object should not be instantiated directly. This class contains some general purpose methods which are useful for any type of gaincurve.

Methods

__call__(*self*, *f*)
Ailased to the gain(f) method

gain(*self*, *f*)
Calling this method directly will throw an error. The user should call the corresponding method in a derived class

`f:` Frequency (MHz). Can be an individual float or numpy array of floats.

Return Value

The relative gain

evaluate(*self*, *span*, *cent*, *resolution*)

Evaluates the whole gaincurve, storing the resulting values in instance variables x and y.

`span:` The total frequency span in MHz over which to evaluate (float)

`cent:` The centre frequency in MHz to evaluate about (float)

`resolution:` The frequency step used to evaluate the span in MHz (float)

find_maximum(*self*, *min*=`None`, *max*=`None`, *start*=`0`)

Finds the frequency corresponding to maximum gain. If upper and lower bounds are specified then a bounded minimisation is used, alternatively an initial value can be specified, in which case an unbounded local minimisation is done.

`min:` Lower bound (MHz) (Note: if specified, max must also be specified) (float)

`max:` Upper bound (MHz) (Note: if specified, max must also be specified) (float)

`start:` Initial guess of the frequency corresponding to maximum gain. Used only if doing an unbounded minimisation. (float)

Return Value

The frequency (MHz) corresponding to maximum gain.

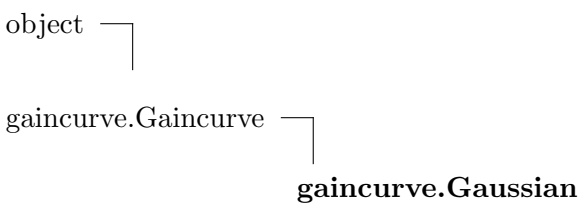
Class Variables

x	Contains the frequency values after evaluate has been called **Value:** `array([], dtype=float64)`

continued on next page

y	Contains the gain values after evaluate has been called **Value:** `array([], dtype=float64)`
cent	Center of the gaincurve in MHz **Value:** `0`
width	Center of the gaincurve in MHz, definition depends on type of gaincurve **Value:** `0`

A.1.4. Class Gaussian



Represents a Gaussian gaincurve for a particular isotope species.

Methods

__init__(*self*, *gas*, *species*, *velocity_offset*=0)

Requires a lasergas object and a species (string) for which to calculate a Gaussian curve.

`gas:`	A Lasergas object, the pressure (for given species) and temperature are taken from here.
`species:`	The name of the species (string) for which to make the Gaussian gaincurve. A species with the given name should be present in the gas mixture.
`velocity_offset:`	If specified moves the curve due to Doppler broadening. Should be a float representing m/s.

Overrides: object.__init__

gain(*self*, *f*)

Returns the normalised gain of the Gaussian gaincurve, centred at the transition of the species (unless a velocity offset was specified).

`f:` Frequency (MHz). Can be an individual float or numpy array of floats.

Return Value

The relative gain

Overrides: gaincurve.Gaincurve.gain

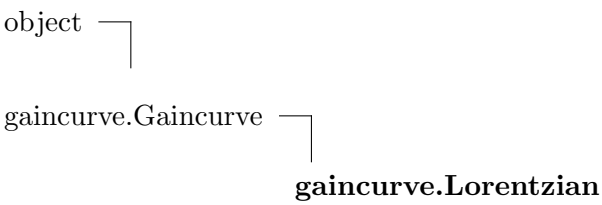
Inherited from gaincurve.Gaincurve(Section A.1.3)

__call__(), evaluate(), find_maximum()

Class Variables

Inherited from gaincurve.Gaincurve (Section A.1.3)
cent, width, x, y

A.1.5. Class Lorentzian



Represents a Lorentzian gaincurve for a particular gas mixture and model of the pressure/collisional broadening.

Methods

__init__(*self*, *gas*, *cent*, *pressure_model*=`'smith'`)
Requires a lasergas object and a species (string) for which to calculate a Lorentzian curve.

`cent:` This variable can be either a string or a float.

If it is a string, it should be the name of a species as defined in physconst module. The constructed

Lorentzian will then have centre frequency on that transition and width defined by that species.

If a float the centre frequency will be that value in MHz and the width will be the FWHM from the weighted average of the species in the gas mixture.

Centre frequency can be reset by changing the instance variable cent.

`gas:` A Lasergas object representing the gas mixture

`pressure_model:` The name (string) of a pressure model known to the physconst module. Defaults to 'smith'

Overrides: object.__init__

gain(*self*, *f*)

Returns the normalised gain of the Lorentzian gaincurve. Centre depends on the value of the instance variable 'cent'

`f:` Frequency (MHz). Can be an individual float or numpy array of floats.

Return Value

The relative gain

Overrides: gaincurve.Gaincurve.gain

Inherited from gaincurve.Gaincurve(Section A.1.3)

__call__(), evaluate(), find_maximum()

Class Variables

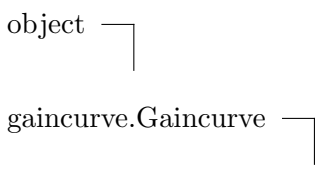
cent — Center of the gaincurve in MHz

Value: `0`

Inherited from gaincurve.Gaincurve (Section A.1.3)

width, x, y

A.1.6. Class Saturation



This gaincurve represents a normalised saturation function, similar to an inverted Lortentzian function.

Methods

__init__(*self*, *gas*, *frequencies*, *intensities*, *isat*=`'smith'`, *pressure_ model*=`'smith'`)

Produces a saturation-function gaincurve for a list of operating frequencies with their associated intensities. Note that the HWHM of the dip is found from the weighted average broadenings of each species in the gas mixture. Note that a ValueError will be thrown if frequencies and intensities are not the same length.

`frequencies:`	list of frequencies (as floats) for the operating modes
`intensities:`	list of intensities (as floats) of the beams corresponding to the list of frequencies.
`isat:`	This can be either the saturation intensity (a number), or alternately it can be a string. If it is a string, it must be the name of a saturation intensity model (see physconst module). In this case the supplied gas mix info will be passed to the given model to determine the saturation intensity. Default model is 'smith'.
`gas:`	A Lasergas object representing the gas mixture
`pressure_model:`	The name (string) of a pressure model known to the physconst module. Defaults to 'smith'

Overrides: object.__init__

gain(*self*, *f*)

Returns the value of the saturation function.

`f:` Frequency (MHz). Can be an individual float or numpy array of floats.

Return Value

The relative gain

Overrides: gaincurve.Gaincurve.gain

Inherited from gaincurve.Gaincurve(Section A.1.3)

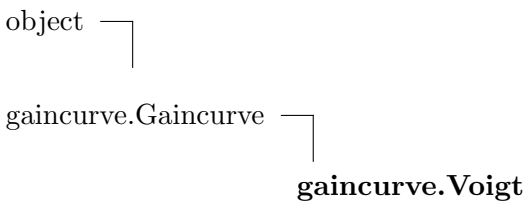
__call__(), evaluate(), find_maximum()

Class Variables

Inherited from gaincurve.Gaincurve (Section A.1.3)

cent, width, x, y

A.1.7. Class Voigt



Defines a gaincurve where the gain at a given frequency is found by detuning a supplied gaincurve (lor) to the desired frequency and convolving this with another supplied gaincurve (gaus).

Integration can be done either continuously (more accurate) or from discretely sampled data (faster when evaluating repeatedly, by an order of magnitude or more). Continuous integration is the default, but this can be changed. See __init__ method documentation for details.

If discrete is supplied then integration will be done discretely using Romberg integration on $2^k + 1$ evenly spaced points. If specifying a min and max value and using discrete integration, be careful to ensure the range given is wide enough.

Note: Romberg integration uses the trapezoid rule at step-sizes related by a power of two and then performs Richardson extrapolation on these estimates to approximate the integral with a higher-degree of accuracy.

Note that the parameters of the constructor gaus and lor are not necessarily Gaussian/Lotentzian gain curves, but often will be. If simulating hole burning, gaus will usually be a Gaussian gaincurve convolved with a Saturation gaincurve.

Methods

__init__(*self*, *gaus*, *lor*, *min*=`None`, *max*=`None`, *discrete*=`None`)

If min and max are not specified then they are taken as 6 standard deviations (from the gaus line width) either side of the transition.

If continuous integration is done, result will be accurate to 0.01%

Note: this will change the offset of the supplied Lorentzian when called.

`gaus:` Underlying Gaussian velocity distribution (gaincurve)

`lor:` Lorentzian 'spectral packet' to use (gaincurve)

`min:` Minimum frequency (MHz) over which to do the integration (float)

`max:` Maximum frequency (MHz) over which to do the integration (float)

Overrides: object.__init__

gain(*self*, *f*)

Does the integration and returns the gain according to the Voigt function. f can be either a float or numpy array of floats and the return type will change accordingly.

`f:` Frequency (MHz). Can be an individual float or numpy array of floats.

Return Value

The relative gain

Overrides: gaincurve.Gaincurve.gain

Inherited from gaincurve.Gaincurve(Section A.1.3)

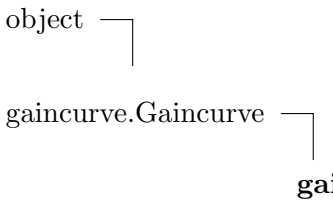
__call__(), evaluate(), find_maximum()

Class Variables

Inherited from gaincurve.Gaincurve (Section A.1.3)

cent, width, x, y

A.1.8. Class Convolve



This gaincurve represents the convolution of multiple supplied gaincurves.

Methods

__init__(*self*, *curves*)

When constructed this curve inherits the maximum values of cent and width from the given curves.

`curves:` a list (or tuple) of all the gaincurves to be convolved.

Overrides: object.__init__

gain(*self*, *f*)

Returns the convolution of the gain from each given curve.

`f:` Frequency (MHz). Can be an individual float or numpy array of floats.

Return Value

The relative gain

Overrides: gaincurve.Gaincurve.gain

Inherited from gaincurve.Gaincurve(Section A.1.3)

__call__(), evaluate(), find_maximum()

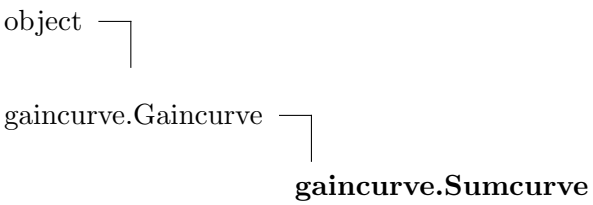
Class Variables

curves	A list of the curves which will be convolved **Value:** `[]`

Inherited from gaincurve.Gaincurve (Section A.1.3)

cent, width, x, y

A.1.9. Class Sumcurve



This gaincurve represents the summation of multiple supplied gaincurves.

Methods

__init__(*self*, *curves*)

When constructed this curve inherits the maximum values of cent and width from the given curves.

`curves:` a list (or tuple) of all the gaincurves to be summed.

Overrides: object.__init__

gain(*self*, *f*)

Returns the gain sum for each given curve.

`f:` Frequency (MHz). Can be an individual float or numpy array of floats.

Return Value

The relative gain

Overrides: gaincurve.Gaincurve.gain

Inherited from gaincurve.Gaincurve(Section A.1.3)

__call__(), evaluate(), find_maximum()

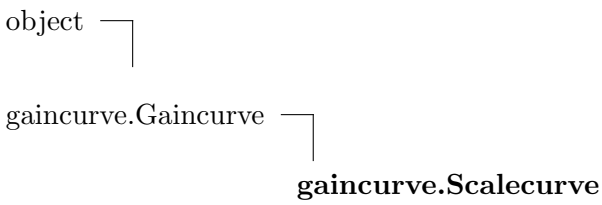
Class Variables

curves	A list of the curves which will be summed **Value:** `[]`

Inherited from gaincurve.Gaincurve (Section A.1.3)

cent, width, x, y

A.1.10. Class Scalecurve



This gaincurve is the equivalent of a scalar multiplier.

Methods

__init__(*self*, *curve*, *m*)

When constructed this curve inherits the values of cent and width from the given curve.

`curve:` the gaincurve to be scaled

`m:` the multiplication factor to use (float)

Overrides: object.__init__

gain(*self*, *f*)

Returns the scaled gain.

`f:` Frequency (MHz). Can be an individual float or numpy array of floats.

Return Value

The relative gain

Overrides: gaincurve.Gaincurve.gain

Inherited from gaincurve.Gaincurve(Section A.1.3)

__call__(), evaluate(), find_maximum()

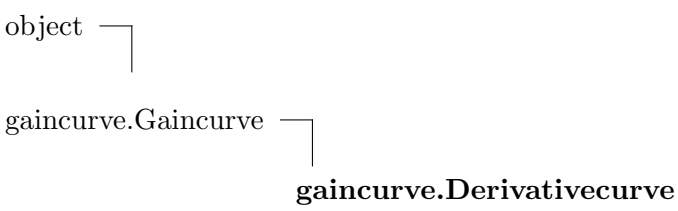
Class Variables

m	The multiplication factor
	Value: `0.0`

Inherited from gaincurve.Gaincurve (Section A.1.3)

cent, width, x, y

A.1.11. Class Derivativecurve



This gaincurve is built from another gaincurve, and represents its derivative at a specified order. This provides a wrapper to scipy.misc.derivative (continuous) or np.diff (discrete). Note that if discrete operation is selected, the resolution parameter has no effect. In this case you need to simply pass an array of frequencies with the desired spacing.

Methods

__init__(*self*, *curve*, *order*=`1`, *resolution*=`1.0`, *discrete*=`False`)
x.__init__(...) initializes x; see x.__class__.__doc__ for signature

`curve:`	The original gaincurve to differentiate (gaincurve)
`order:`	The order of the derivative, defaults to 1st order (int)
`resolution:`	The resolution over which to calculate the derivative, defaults to 1 MHz (float)

Overrides: object.__init__

gain(*self*, *f*)

Returns the derivative of the curve at the given frequency.

`f:` Frequency (MHz). Can be an individual float or numpy array of floats.

Return Value

The relative gain

Overrides: gaincurve.Gaincurve.gain

Inherited from gaincurve.Gaincurve(Section A.1.3)

__call__(), evaluate(), find_maximum()

Class Variables

Inherited from gaincurve.Gaincurve (Section A.1.3)

cent, width, x, y

A.1.12. Class LaserGas

object

gaincurve.LaserGas

Methods

__init__(*self*, *temperature*, *transition*)

x.__init__(...) initializes x; see x.__class__.__doc__ for signature

`temperature:` gas temperature in K (float)

`transition:` the name (string) of a transition in spectral notation with a -> separating the initial and final state. Must be known by the physconst module.

Overrides: object.__init__

add_species(*self*, *name*, *partial_pressure*, *lasing*)

Adds a new species to the gas mixture

`name:` The name of the species (string). Must be known to the physconst module

`partial_pressure:` The partial pressure (float representing mbar) of the species

`lasing:` Boolean indicating if the species is lasing.

get_ratio(*self*, *name*)

`name:` name of the gas species to get ratio for (string).

Return Value

The fraction that the named gas species represents of the total gas pressure.

get_lasing_ratio(*self*, *name*)

This is essentially the same as get_ratio, but only applies to species which are known to be lasing

pressure_broadening_average(*self*, *pressure_model*=`'smith'`)

Caclulates the FWHM of pressure broadening for the total pressure of the gas mix but with the model weighted by the partial pressures of the lasing gas species.

`pressure_model:` Which model of the pressure/collisional broadening to use. Must be a name (string) known by physconst module. Defaults to 'smith'

Return Value

The full width at half max (in MHz) of the pressure broadening

make_composite(*self*, *iisat*=`0`, *hole_separation*=`0`, *unsat_max*=`'auto'`, *discrete*=`None`, *debug*=`False`)

Returns a composite gaincurve for this gas mixture, found by constructing and weighting appropriately Voigt curves for each lasing species. The returned gaincurve will be normalised to the gain value at the unsaturated maximum.

If the iisat parameter is specified, holes are burned opposite the unsaturated maximum and separated according to hole_separation parameter. Note that

the frequency of the unsaturated maximum is stored in the class instance variable unsat_max if it is not known. It is then used if later calculations. If the gas mixture is changed meanwhile, you will have to run this method forcing the recalculation of a new unsat_max.

`iisat:`	Fractional intensity / saturation intensity (assumed to be the same for both holes)
`hole_separation:`	Frequency difference between the holes in MHz
`unsat_max:`	A string describing what to do regarding calculation of the unsaturated maximuim. 'auto' will calculate it if unknown but use the saved value otherwise, 'recalculare' will force a recalculation, 'disable' will do nothing.
`debug:`	Print debugging/progress information (boolean)

count_crossings(*self*, *iisat*, *fsr*=`50`)

This method returns the number of positive (i.e., -vt to +pv) zero crossings in the derivative of the composite curve for a gas mixture.

find_threshold(*self*, *fsr*=`50`, *plt*=`None`)

Returns the iisat value at multimode threshold, assuming the definition of the multimode threshold is the lowest value of iisat for which their exist 2 roots in the 1st derivative of the gain curve with a positive 2nd derivative.

This is a simple uphill-only optimisation routine which find the point when the number of crossings = 2. Works by stepping up the isat, until we get to 2 crossings, then backtracking 1 step and increasing the precision.

Class Variables

unsat_max	Frequency corresponding to the maximum of the unsaturated gain curve. **Value:**

Instance Variables

temperature	Mean gas temperature (K)
transition	Transition identifier string
species	Dictionary where keys are the element names (strings) and values are the partial pressures (mbar) (float)
lasing	Dictionary where keys are the element names (strings) and values booleans indicating if that species is lasing on the given transition.
pressure	Total pressure. This is automatically updated. Don't change this manually.

Implementation: gaincurve.py

A.2. Physconst module

A.2.1. Synopsis

This module is used as a repository for a variety of physical constants and phenomenological numerical models.

A.2.2. Functions

mass(*thing*)

Returns the mass in kg of something. If the thing is unknown, then a ValueError will be thrown and the names of the available things returned.

`thing:` name (string) of thing to find the mass of. Currently available 'Ne 20', 'Ne 21', 'Ne 22', 'He 4', 'earth'

natural_abundance(*thing*)

Returns the natural abundance fraction of the given thing. Note that if the thing is unknown, then a ValueError will be thrown and the names of the available things returned.

`thing:` name (string) of thing to find abundance of, currently available

'Ne 20', 'Ne 21', 'Ne 22'

transition_frequency(*species*, *transition*)

Returns the transition frequency in MHz for a given transition of a given species. Note: A list of known state names and species will be returned in the ValueError information if given incorrect of empty names.

`species:` Species name as a string, e.g. 'Ne 20'

`transition:` The transition is encoded with a -> used to separate states.

lorenzian_linewidth(*gas*, *model*=`'smith'`, *uncertainty*=`False`)

Returns the Lorentzian line width (FWHM) in MHz for a given gas mixture according to a supplied model. Note that a ValueError will be raised if the species or model could ot be found.

`gas:` LaserGas object, must have pressure and temperature specified.

`model:` Name of species, currently available: 'smith', 'tuchin', 'bennett'

`uncertainty:` If true, function will return a tuple with width and the width error if it is available for the model. Otherwise will return the width by itself.

saturation_intensity(*gas*, *model*)

Returns the saturation intensity for a given LaserGas object using the specified model in W/cm^2. A ValueError will be raised if the specified gas parameters are out of range of the desired model.

`gas:` LaserGas object, must have pressure and temperature specified.

`model:` Name of the model to use (string). Available models are 'smith', 'graham'.

A.2.3. Variables

k — boltzmans constant (m^2 kg/s^2/K)
Value: `1.3806503e-23`

continued on next page

amu	atomic mass unit (kg)
	Value: `1.66053886e-27`
c	speed of light (m/s)
	Value: `299792458.0`
R	Gas constant (J/K/mol)
	Value: `8.314472`
h	Plank constant m^2 kg / s
	Value: `6.6260683e-34`
N_A	Avagadro's number
	Value: `6.0221415e+23`

Implementation: physconst.py

A.3. Pysolid module

A.3.1. Synopsis

This module was written to provide an easy to use programming interface to the Fortran 77 program by Dennis Milbert. This is used for calculation of displacements due to solid earth tides through an implementation of section 7.1.2 of the IERS Conventions (2003). Pysolid extends this functionality to calculate stains, tilts and rotations induced by earth tides.

All of these calculations require finding the coordinates of test points relative to the observatory. Vincenty's 'direct' formulae is used for this. Pysolid supports reference ellipsoids WGS 84, GRS 80 and WGRS 80/84 however in all the work presented in this thesis the default WGS 84 is assumed.

This depends on the scitime module, which automatically maintains a table of leap-second data (automatically downloading updates from IERRSS when necessary and conversion routines necessary for the various astronomical calculations.

A.3.2. Functions

gps_time_array(*s*)
Convienence function : Returns a array of integers in the form year, month, day, hour, min, sec in GPS time given a scitime object.

isLatitude(*latitude*)
`latitude:` Valid latitude in decimal degrees

Raises

`ValueError` If not a valid latitude

isLongitude(*longitude*)
`longitude:` Valid longitude in decimal degrees

Raises

`ValueError` If not a valid longitude

isAzimuth(*azimuth*)
`azimuth:` Valid azimuth in decimal degrees

Raises

`ValueError` Outside the range 0..360

vinc_pt(*latitude*, *longitude*, *azimuth*, *distance*, *reference_ellipsoid*=`'WGS 84'`)
Implementation of Vincenty's Direct formulae.

`latitude:`	Latitude of the initial point in decimal degrees
`longitude:`	Longitude of the initial point in decimal degrees
`azimuth:`	Angle in degrees from the initial point in decimal degrees
`distance:`	Distance from the initial point to the new point in meters
`reference_ellipsoid:`	Reference ellipsoid to be used in coordinate calculations. May be specified, either 'WGS 84' (default), 'GRS 80' or 'WGRS 80/84'.

Return Value

Projected point and reverse azimuth

(type=(phi2, lambda2, alpha21) as a tuple)

Raises

`ValueError` in event of incorrect reference ellipsoid or invalid longitude or latitude

Author: Robert Hetland. Modified from code taken from PyROMS suite.

License: BSD license (2007)

Note: the inverse formulae may give no solution over a line between two nearly antipodal points. This will occur when lembda is greater than pi in absolute value. (Vincenty, 1975)

A.3.3. Class Observatory

Methods

__init__(*self*, *latitude*, *longitude*)
Initialises an observatory at the given location.

`latitude:` latitude of the site in decimal degrees

`longitude:` longitude of the site in decimal degrees

setLatitude(*self*, *latitude*)
`latitude:` latitude of the site in decimal degrees

Raises

`ValueError` if latitude is invalid

setLongitude(*self*, *longitude*)
`longitude:` longitude of the site in decimal degrees

Raises

`ValueError` if longitude invalid

getDisplacement(*self*, *starttime*, *step*, *duration*, *offset*=`(0, 0)`)
`starttime:` Time to begin calculation for

`step:` Time between calculated sample in seconds

`offset:` Point relative to this observatory to calculate displacement for. Represented as a tuple containing the distance in m and the angle in degrees from the observatory.

`duration:` Total number of seconds to calculate for.

Return Value

An array of displacements in meters x,y,z calculated at this observatory, or a given distance away if offset is specified.

getReferencePoints(*self*, *starttime*, *step*, *duration*, *angle*, *baselength*, *numpoints*)

`starttime:` Scitime object representing start time of displacement data

`step:` Time step in seconds to use

`duration:` Total duration in seconds

`angle:` Angle in degrees of the first point

`baselength:` 2* the distance in meters from the observatory to each point

`numpoints:` Number of points to be equally spaced around a full circle

Return Value

List of the displacements for n reference points centred about the observatory

Note: This methods is normally used internally by getStrain etc

getStrain(*self*, *starttime*, *step*, *duration*, *angle*, *baselength*=200000)

`angle:` angle in decimal degrees along which to find strain

`baselength:` length of the test vector in m.

Return Value

Strain (as a fractional extension) along a given vector

(type=`numpy.array`)

Note: Other parameters the same as in `getDisplacement`

getRotation(*self*, *starttime*, *step*, *duration*, *baselength*=200000)
`baselength:` length of the test vector in m.

Return Value

Rotation angle in degrees along a given vector

*(type=**numpy.array**)*

Note: Other parameters the same as in `getDisplacement`

getRotationRate(*self*, *starttime*, *step*, *duration*, *angle*, *baselength*=200000)
Return Value

rotation rate in degrees/sec of test vector.

*(type=**numpy.array**)*

Note: see `getRotation` for other parameter definitions

getTilt(*self*, *starttime*, *step*, *duration*, *angle*, *baselength*=200000)
Finds the earth tide induced tilt by defining a plane around the observatory and calculating the normal to the plane defined by a sample triangle centered at the observatory.

`angle:` Angle in degrees as on the observatory compass to calculate the tilt for. For example, north would be 0.

Return Value

Tilt angle in degrees from the given vector projected onto the horizontal plane, i.e. nominally 90. For example if the given angle is 0 (pointing north) then a returned value of 89 indicates a tilt of 1 degree to the north.

Note: Other parameters are the same as in `getDisplacement`

Class Variables

longitude	Longitude of the observatory in degrees. Do not set manually. Use `setLongitude` method. **Value:**
latitude	Latitude of the observatory in degrees. Do not set manually. Use `setLatitude` method. **Value:**

Implementation: pysolid.py

B. Data Acquisition Software

B.1. Troubleshooting

This section contains a series of operational notes for troubleshooting various problems that can arise with the various data acquisition and storage programs that have been put into operation by the author.

B.1.1. Database Problems

Problem: Database system is indicating that some files are showing checksum errors or can not be read.
Check: The network connection to and from (check both directions) the system generating the data. When the network problems are resolved, the data should be manually copied from the DAQ systems archive.
Related: It is possible for `.md5` files to accumulate in the `/data/incoming` directory if no corresponding `.xml` files have been copied. This is because the `organise_incoming.py` script only checks for the `.xml` files, then finds the corresponding `.md5` file. Since `.md5` files are much smaller, they are more likely to be successfully copied in the event of network problems. In this case their existence gives a handy indication of which `.xml` files need to manually be copied once the network problems have been resolved.

Problem: Database queries are performing particularly poorly.
Check: The data in the database. Particularly large documents with very many XML tags are likely to cause this problem. For example, omitting the bindata attribute on high speed measurements or setting a small number of blocks to average

for non-bindata measurements will cause this problem.

Problem: Data is accumulating in `/data/incoming` and not being added to the database. Running the `organise_incoming.py` script gives the message 'another process is already organising data', although `ps -ax | grep organise_incoming` reveals no such process is running.
Check: The file `/tmp/organising_progress`. This is written by the `organise_incoming.py` script and contains one line with the command currently being run. It should be deleted once the script has finished, although if it is not for some reason then further runs of the script will be aborted. Delete this file to resolve this issue. A look inside the file should reveal the last command successfully run by the script which may give some insight as to the original problem.

Problem: You have some data showing up in queries that is unexpected in some way. You want to find the document in the database which contains this particular record.
Solution: Write a simple xquery to obtain the record you are looking for, say `$a`. Then use `document-uri(root($a))` to obtain the URI of the document that contains that record. You can then go and investigate it manually.

B.1.2. DAQ System Problems

Problem: DAQ system has been failing the specified sanity checks, although it was later determined that the data is usable and the sanity check conditions were set to finely. It is necessary to manually recover a portion of data from the suspect measurements file, which has grown very large.
Solution: Use the `build_xml.py` script as you normally would, but use the `-t` flag, followed by the time range you want to recover, as well as the `-e` flag. Note that there exists a simplified version of the `build_xml.py` script which does not have this functionality; to check the functionality and for more information on how to use the program, pass the `-h` flag.

Problem: No data is being acquired, or not on some channels
Check:

- The Device ID and all other controls in DAQ setup tab
- Has a sanity check condition has been defined for each channel? Note that you do not have to specify any sanity checking conditions (i.e. check DC, RMS and Freq can be false) but there must be an entry in the array of sanity check conditions for each channel index used, i.e. they can't be ghosted out if the channel is being used, even if no checks are actually being applied.
- Can waveforms be acquired manually with `acquire_single_waveform.vi` ?
- Is the input terminal configuration set correctly for the breakout box you are using? Breakout boxes should all be wired in differential mode, though some older devices are not. Check the code in the measurement task setup section.
- Is the driver for the DAQ card working? Remember, the driver must be compiled for the version of the kernel you are running. If you have updated the kernel version using an update tool then you will have to updated the driver, following NI's instructions. They provide a simple script `updatenidrivers` which you can use for this, though note that you will need the full kernel sources installed for the kernel you are using. Alternatively, you can just go back to the old kernel by selecting it from the GRUB bootup menu.

Problem: Data is not being copied from DAQ system to the database, and is accumulating in the archive directory of the DAQ system.
Check: Network connection, manually copy files to `/data/incoming` on database server. If using scp to copy the data, try connecting via ssh to the database server. A password should *not* be required. If a password is required, then ssh-agent might not be working. ssh-agent essentially 'remembers' your ssh passwords for you. Many tutorials can be found for setting up ssh-agent, but broadly it involves making a key on the client DAQ system by running `ssh-keygen -t -rsa` while in the users `/.ssh` directory, then appending the contents of the resulting .pub file to the `/.ssh/authorized_keys` file on the server. Further problems can be diagnosed

by passing the `-v` flag to ssh to see what is going on.

Problem: NTP is no longer locking to the correct server
Check: The status of the NTP synchronisation can be checked by running `/usr/sbin/ntpq -p`. Times are in msec. The server locked to should have a `*` in the first column.
Related: Standard red-hat systems are set by default to update NTP on bootup. If setting up a new system, you may have changed the ntp.conf or other NTP settings correctley, but they may be overwritten the next time the system reboots. This behaviour can be disabled in `/etc/sysconfig/network-scripts/ifcfg-eth0` (assuming eth0 is the default network device).

Note: at the time of writing, the primamry ntp server at Canterbury should resolve to `linux3.cosc.canterbury.ac.nz`

Problem: After making code changes, the start and stop buttons are not viable.
Check: The size of the front panel window probably changed, cutting off the buttons at the bottom. With some screen resolutions, LabVIEW sometimes changes the front panel window size when editing. Make sure when you have finished editing to enlarge the front panel window so that it is big enough to show all buttons.

Problem: You have some timing information in Unix time stamp format and you want to check the civil time which this corresponds to.
Solution: Most programming languages have libraries for parsing dates and times which can be used for this, however there is also an easy way to do this directly from the command prompt on any Unix computer. For example, `date --utc -d @1282186108` will tell you the civil time corresponding to the given (UTC) time stamp. For local time, omit the `--utc`. Also note that `date +%s` will tell you the current time stamp.

Problem: When connected to a DAQ machine via SSH (console only) you want a quick way to check that data is being acquired and see the values.
Solution: As the data being acquired is continually appended to the `elements` file in the directory corresponding to the respective sensor, is it convenient to simply view the last line in this file. This can be done with `tail -1 elements` in the sensor directory, for example `/mnt/data/xmldata/rldata/instrument!g0/sensor!Sagnac`. If you would like to monitor the data on the console the `watch` command is useful, for example `watch -n 0.2 tail -1 elements` will show you the data being acquired, updating every 0.2 s.

B.2. Scripts

B.2.1. universal_timeseries.py

Synopsis

This program obtains a simple time series from the ring laser database in a variety of formats. It works transparently with both data in binary series elements or regular `<measurement>` blocks.

Notes

- This program can also be imported as a python library, used via the command line or via the web interface.
- Only works with time in UTC-Unix format, MJD not supported
- The `-u` option is intended to speed up queries for quick measurements. If used with measurements encoding binary series data, note that it applies to *each block* of measurements so the output is likely to have non-uniform gaps. Use the slower -r option instead if you want to avoid this.
- `png` or `pdf` output options produce a plot.
- `show` output option will start an interactive plot window.
- `mat4` output option produces a array compatible with MATLAB 4 or octave.

- `mat5` output option produces a array compatible with MATLAB 5 or later.
- Re sampling/realignment not applicable if spline interpolation selected.

Usage

```
./universal_timeseries.py [options]
```

Options (compulsory)

-o	[csv | npy | mat4 | mat5 | pdf | png | show]	output format to use
-i	instrument	name of instrument
-s	sensor	name of sensor
-b	year,month,day,hour,min,sec	begin time
-e	year,month,day,hour,min,sec	end time

Optional

-h		display this help information
-f	filename	name of output file (no extension)
		(compulsory unless using -o show)
-d		show debugging information
-c	server:port	connect to this server
		defaults to physring:8080
-z		realign time stamp to begin at 0 in the output data
-g		output only the data up to the first gap
-u	N	under sample retrieved data, obtaining
		one in every N measurements
-r	period	re sample period (sec)
-a		when used with `-r`, will align samples to
		whole UTC seconds.
-v	[exclude | include | only]	*include* measurements enclosed in event
		structures as well as regular
		measurements, or output *only* data.
		in events. Default behaviour is to *exclude*
		all data in event structures.

Example

```
./universal_timeseries.py -o csv -i ug3 -s Sagnac -b 2010,1,1,1,1,1 -e
2010,1,2,1,1,1
```

Implementation: universal_timeseries.py

B.2.2. buildxml.py

Purpose

Builds up an XML file given a root path, assumed to be created by LabVIEW rlxml VIs

Usage

```
./buildxml [options] rootpath
```

Options

-h		help
-d		debug
-c		compress
-m		md5 checksum
-v	filename.xml	validate against given schema (relaxNG supported)
-o	filename	output filename (optional, defaults to standard a value)
-r		remove rootpath after successful run

Implementation: buildxml.py

B.2.3. binary_series.py

Synopsis

This program takes a ring laser XML data file and identifies any series of measurements inside numeric sensors which have the 'bindata' attribute. These are replaced by a single measurement with `<binaryseries>` elements encoding the same data as base64 numpy arrays. This vastly reduces file size and the number of elements so the data can be put into a database or other structure with minimal performance penalty.

Notes

The 'bindata' attribute should be float32, float64 (default) or float128. This program uses ElementTree objects so will use at least as much RAM as the file size. Source file can be `.xml` or `.xml.bz2` or `.xml.zip` Note that if a `.zip` is used, it is assumed to contain only one file.

Usage

```
./binary_timeseries.py [options] file.xml
```

Options

-h		help
-d		debug
-m	file.md5	check source file against md5 hash file
-v	schemafile.rng	validate result against RELAXNG schema file
-o	outputfile.xml	processed file to write, if omitted original will be replaced.

Implementation: binary_series.py

B.2.4. xmldbmerge.py

Synopsis

This program takes a .xml file and merges it to an eXist-db database.

If the filename contains a block of numbers separated by hyphens (the block should be delimited by underscores) then the file will be merged to a sub collection defined by the numbers in reverse order. E.g. if root is `/db/rldata` and file name is `db-ug3_21-12-2009_1028.xml` file will be merged to: `/db/rldata/2009/12/21/db-ug3_21-12-2009_1028.xml`

Notes

If the file name does not conform to this format then the file will be merged to the 'unknown' collection in the root path.

Usage

```
./xmldbmerge.py [options] file.xml
```

Options

-s	host:port	eXist-db server to connect to. Defaults to localhost:8080
-u	username	Username to connect with
-p	password	Password for the given user
-r	/rootpath	Path of the root collection
-h		This help information

Implementation: xmldbmerge.py

B.2.5. reconcile_quakes.py

Synopsis

This program queries the ring laser XML database for recent events and then reconciles them with earthquakes available on the geonet web service. It then builds an XML file containing a corresponding eventinfo element and saves it to a specified location.

At present this works just with geonet, but the quake service inherits from a common interface, so it should be easy to write a new class inheriting `quake_service` for a different service.

It is possible to run this program from the command line or import it as a python module and use it to access the quake services.

Usage

```
./reconcile_quakes [options]
```

Options

-s, –server	host:port, host:port	eXist-db server to connect to. Defaults to localhost:8080
-p, –path	path	Where to put the eventinfo file. File name will follow the form: `eventinfo_21-12-2009_1028.xml`
-b, –daysback	days	Number of days back to try to reconcile. Defaults to 10
-x, –proxy	user:pass@proxy:port	Proxy server to use to access internet Example: `http://user:pass@webgate.canterbury.ac.nz:3128`
-h, –help		This help information

Implementation: reconcile_quakes.py

B.2.6. autocovar.py

Synopsis

This program computes the autocovariance (the un normalised autocorrelation) of Sagnac data on a given instrument using data retrieved from the ringlaser eXist-db server.

Notes

- Works only with data encoded as bindata.
- Requires the nominal scale factor to be defined for the instrument.
- Minimum period is fixed at 1.5 s.
- All elements of start and end times should be integers.
- Does not include quake events by default.

Usage

```
./autocovar.py [options]
```

Options (compulsory)

-i	instrument	name of instrument
-s	year,month,day,hour,minute	starttime
-e	year,month,day,hour,minute	endtime

Options (optional)

-h		this information
-d		display plot
-m		compute the exact period at maximum
-t		text output
-p	period	maximum period (default 50)
-r	period	step/resolution in period (default 0.2)
-f	filename	save output to a `npz` file

Example

```
./autocovar.py -i ug3 -s 2010,1,1,1,1 -e 2010,1,2,1,1 -p 200 -r 0.2 -m
-t
```

Implementation: autocovar.py

B.3. Ring Laser Data XML schema

The complete specification for the ring laser XML format in the RelaxNG (full XML syntax) is available here[1]. The 'natural language' equivalent to this is presented in Section 6.5, the formal specification is made available here to serve as a precise reference.

Implementation: rlxmlspec_4_2010.rng

B.4. Ring Laser LabVIEW VI package

In order to simplify the creation of data acquisition software and maximise re-use, a standard 'toolbox' of LabVIEW VIs has been created. These are distributed in a `loging` directory containing subdirectories mirroring the sections in the following documentation. The distribution at the time of writing is available here (attached to this page). Note that this does not include some utility VIs from the OpenG distribution[2].

An example data acquisition program `loger_template.vi` is included. This can

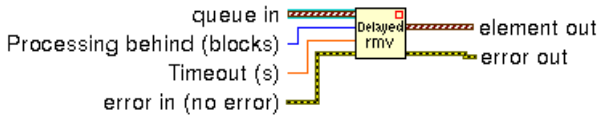
[1] Double click the link text or use your PDF readers file attachment toolbar.
[2] http://wiki.openg.org/Main_Page

serve as a template for building a data acquisition program. It already has all the user interface elements such as tabs and monitor controls, loging, as well as the producer-consumer design implemented. Normally the user would only have to modify the inner consumer loop, the large outer loop is really only for setting up the user interface and responding to events.

Implementation: loging.tar.bz2

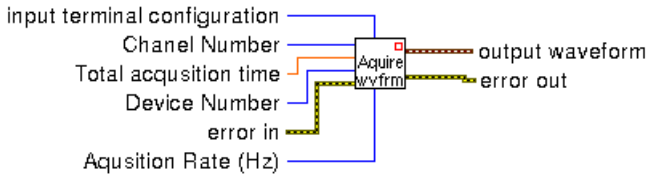
B.4.1. Data Acquisition

queue_delayed_remove.vi



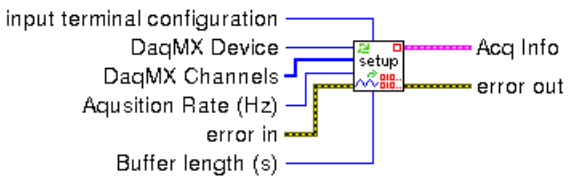
This VI is a drop-in replacement for the standard LabVIEW remove from queue VI. The only exception is that it has an additional input, *processing behind (blocks)* which requires an integer. This VI will only return an item from the queue provided that the queue has more than the given number of items, the effect is to delay the removal. This is useful delaying the processing of raw data and ensuring that the raw waveform data beginning a few second before a seismic event can be saved.

acquire_single_waveform.vi



This VI handles everything necessary to get a single waveform of finite length from a specified data acquisition device and frees any resources when the task is finished. It is useful to run this before starting an acquisition (so any errors with the hardware will show up at this early stage) or for obtaining a waveform from a single channel in a simple program.

setup_multi_channel_acquisition.vi



This VI takes care of setting up and starting a continuous acquisition on a group of channels (passed in as as an array) for a specified data acquisition device. The buffer length defaults to 10 s but longer lengths can be specified if required.

Output *Acq Info* (acquisition information) is a cluster containing the start time (LabVIEW time stamp cast as double), rate per channel and the DAQ Task Constant (cast as a variant). This provides a consistent interface to acquisition information and allows different drivers (which have different underlying task data types) to be used from a consistent interface.

acquire_multi_channel.vi

This VI is used for acquisition over multiple channels. It automatically builds an array of waveform data (one entry for each channel) and samples the computer clock each time it is called in order to ensure the gathered waveforms are not subject to clock drift. This should be used with `setup_multi_channel_acquisition.vi`.

close_multi_channel.vi

Given an acquisition task (described by the acquisition information cluster), this stops the task and frees any resources. This is intended to be used with a task information cluster generated by `setup_multi_channel_acquisition.vi`.

B.4.2. User Interface

monitor.vi

This polymorphic VI allows individual data points or waveforms to be displayed on a central 'monitor' waveform graph. A waveform or data point (typically a double) can be wired to this VI. A descriptive name (string) also is required. For each instance of this VI, the given name will be added to a central drop-down box. The user can then select which data to display on the graph.

The waveform graph and associated combo box must already exist and refnums to them have been set to a global variable `monitor_widgets_global`.

monitor_widgets_global.vi

This global variable is a cluster which contains refunms to a global graph and other widgets which are used to display data being monitored by the monitor VI(s). You need to create refnums for the various widgets and wire them to the inputs of this cluster. You need a 'Waveform Graph', 'Combo Box' (for selecting item to monitor), 'Update' boolean/check box (for controlling updating of the display) and 'Points to plot' (integer, controls number of points to display on the graph).

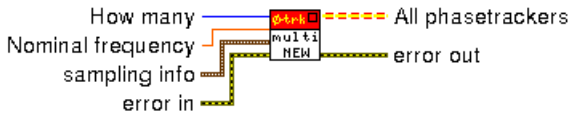
B.4.3. Phase tracking

Phase tracking is implemented with a LabVIEW class `phasetracker.lvclass`. Phase tracker objects are recognisable by their yellow and red wires.

phasetracker.lvclass:new_phasetracker.vi

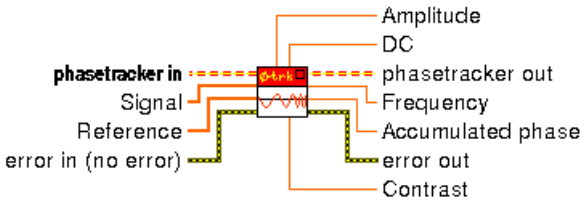
Creates a new phasetracker object. The kernel is created at this stage, so this requires a nominal frequency to be specified as well as a sampling information cluster. The sampling information contains the sample rate and number of samples to acquire in a block.

phasetracker.lvclass:multi_new_phasetracker.vi



This is a useful VI which creates an array of identical phase tracker objects. This is useful when tracking the phase of multiple signals together.

phasetracker.lvclass:track_phase.vi

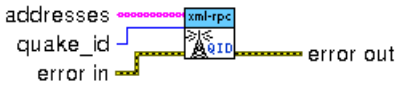


This VI implements the phase tracking. When called, it requires arrays for one block of the waveform being tracked and the reference waveform (with frequency already specified when the phase tracker was created). The outputs available indicate the accumulated phase, frequency, amplitude (RMS), DC level and contrast (AC/DC) ratio.

B.4.4. Notification

The notification VIs make use of the LabVIEW XMLRPC server/client (by Mark E Smith). It is required that this be installed in an XML-RPC subdirectory along side the notification VIs. The XML-RPC server should of course be running on any system you wish to use these VIs with.

send_quakeid.vi



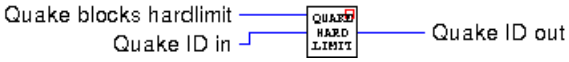
This VI is responsible for notifying one or more systems of a quake by calling the `quakeid_listener.vi` method over XMLRPC listening on port 8080 (default). A quake ID must be passed, which is normally the time stamp corresponding to the first occurrence. Note that a quake ID of -1 indicates no quake. It is necessary to send a -1 ID once the quake has finished.

Systems are identified by network name (or IP address) in the array of strings 'addresses'. Higher priority systems should be higher on the list.

quakeid_listener.vi

This VI must be located somewhere accessible to the XML-RPC server, for example the `XML-RPC/XML-RPC Server/XML-RPC Methods` directory. This VI is called remotely over XML-RPC and responds by setting the global `quake_id_global` to the given quakeID.

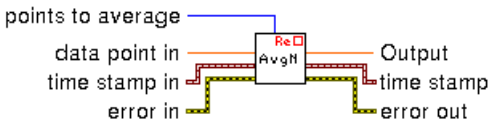
quake_hard_limit.vi



This reentrant VI can be used to set a hard limit on the number of blocks processed for a quake. It is a useful safeguard if for some reason an event is not sent notifying the system of the end of a quake. This works by checking a quake ID, and if a specified number of consecutive identical quake IDs are passed it will output -1, i.e. no quake. Otherwise, the input quake ID is passed through.

B.4.5. Processing

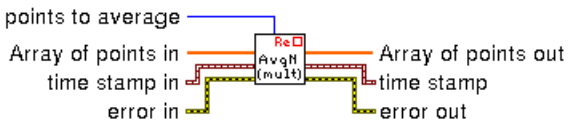
point_averager.vi



This reentrant VI is used for averaging a series of data points with associated time stamps. It works by accumulating the given number of values, then outputting the

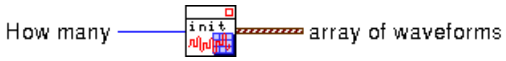
average value and time once the given number has been accumulated. In the cases when data is accumulating and the output is not ready, a `NaN` value will be output. The following code should be able to accommodate this. This can be used with the RL XML VIs, these will not save any data if given a `NaN` input.

multipoint_averager.vi



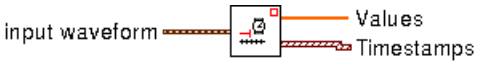
This is a version of `point_averager.vi` which works with whole arrays of points and can thus be used over multiple channels.

init_wvform_array.vi



This is a useful utility VI for initialising an array of a given number of empty waveforms.

waveform_to_arrays.vi



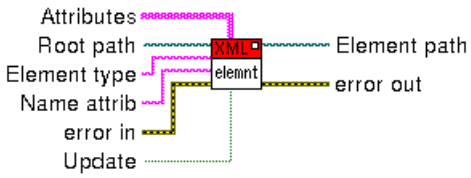
This is a simple utility VI that takes a waveform and obtains arrays of the values and timestamps.

wvfm_unixtime.vi

This is a simple utility VI which takes a waveform and returns the time in the middle of the waveform as a UNIX time stamp.

B.4.6. Ring Laser XML

rlxml_element.vi



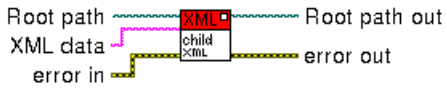
This VI is for creating a new XML element which will then be populated with child elements.

It works by creating a directory (and if necessary) an *attributes* file in a directory structure intended for flattening into an XML file by the `buildxml.py` script.

The attributes input requires a 2D string array containing the attribute name (first column) and value (2nd column).

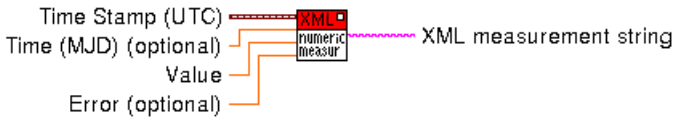
A boolean setting *update* is available. This defaults to true, and for normal measurements should be used in this configuration. However, if set to false the VI will not update the directory structure or elements file. It will only output the new path for sub-elements/data to be written. This can be used for performance reasons, for example when writing out a lot of measurements at high speed, where disk IO would be an issue. Disable updating only with care; make sure at least one iteration has occurred to set up the directory structure before disabling updating. Be sure not to delete the directory structure while updating is disabled.

rlxml_childxml.vi



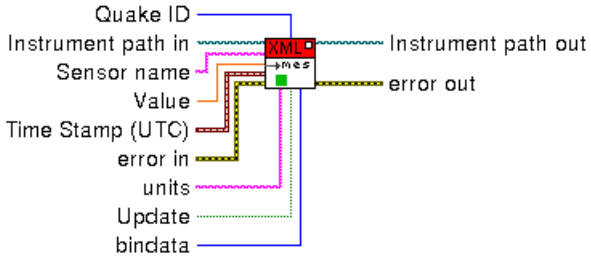
This VI is for writing some XML elements to file. It requires a string (the valid XML data) and if non-empty, appends the string to an elements file in the specified root path. This file would normally be in a directory structure intended for flattening to an XML file by the `buildxml.py` script.

rlxml_numeric_measurement.vi



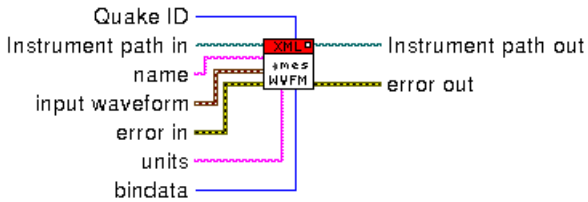
This VI creates an XML `<measurement>` element following the ring laser specification. This works for a numeric, time domain measurement compatible with the ring laser XML data standard. The given time stamp (a LabVIEW timestamp) will be converted into UTC Unix time.

mes_shortcut.vi



This VI provides a shortcut to making a numeric measurement and applying it to a numeric sensor using the most common options. It will also read the global `quake_id` and if a quake is occurring then the measurement will be encapsulated in a `<event>` element with name attribute `eq` followed by the quake ID.

wvfm_mes_shortcut.vi



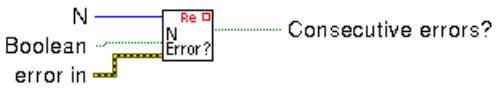
This is the equivalent of `mes_shortcut.vi`, but works with whole waveforms.

sep.vi

This global contains the file path separator used on the platform. You will need to change this if you move the application to a different platform. For example, on linux `/` is used, on windows \ is used.

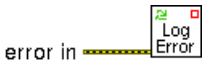
B.4.7. Error checking

consecutive_errors.vi



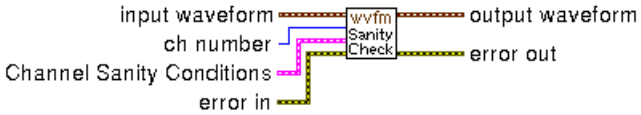
This reentrant VI takes an error cluster and outputs true if N consecutive errors have been passed to it. This is useful for ignoring 'transient' errors. An error is assumed if either the error cluster input indicates an error or the boolean input is true.

log_error.vi



If an error cluster indicating an error is passed in, the error details will be saved to a global log file (given by a global file refnum, `error_log_path.vi`) along with a time stamp.

waveform_sanity_check.vi



This VI checks a waveform to find if the DC, RMS and frequency fall within specified 'valid' ranges. The valid ranges are specified by a sanity check cluster. If the given data falls out-of-range, an error condition will be set in the output error cluster. Use of this VI is recommended in the consumer loop before any data processing is done.

List of Figures

List of Tables

List of Program Listings

Bibliography

[1] Schreiber K, Velikoseltsev A, Rothacher M, Klugel T, Stedman G and Wiltshire D 2004 *Journal of Geophysical Research* **109**

[2] Pancha A, Webb T, Stedman G, McLeod D and Schreiber K 2000 *Geophys. Res. Lett* **27** 3553–3556

[3] Sagnac G 1913 *Comptes Rendus Acad* **157** 708

[4] Stedman G E 1997 *Rep. Prog. Phys* **60** 615–688

[5] Michelson A and Gale H 1925 *Astrophysical Journal* **61** 140

[6] Michelson A and Gale H 1925 *Nature* **115** 566

[7] Macek W, Davis Jr D, Olthuis R, Schneider J and White G 1964 *Quantum Electronics*

[8] Hurst R B, Wells J P R and Stedman G E 2007 *Journal of Optics A: Pure and Applied Optics* **9** 838

[9] Jacobs S F and Zanoni R 1982 *American Journal of Physics* **50** 659–660

[10] Currie B E, Stedman G E and Dunn R W 2002 *Applied Optics* **41** 1689–1697

[11] Bowden G Presented at 8th International Workshop on Accelerator Alignment (IWAA 2004), Geneva, Switzerland, 4-7 Oct 2004

[12] Aronowitz F and Ross M 1971 *Laser applications* (Academic Press New York)

[13] Podgorski T and Aronowitz F 1968 *Quantum Electronics, IEEE Journal of* **4** 11–18

[14] Aronowitz F 1999 *RTO AGARDograph (Optical Gyros and their Applications) Part 3* vol 339 (North Atlantic Treaty Orginization)

[15] Aronowitz F and Lim W 1978 Dispersion and gas flow effects in the ring laser gyro *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series* (*Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference* vol 157) ed Ezekiel S and Knausenberger G E pp 7–12

[16] Podgorski T and Aronowitz F 1967 *Quantum Electronics, IEEE Journal of* **3** 235–236

[17] Chow W, Hambenne J, Hutchings T, Sanders V, Sargent M and Scully M 1980 *Quantum Electronics, IEEE Journal of* **16** 918–936

[18] Tobias I and Wallace R 1964 *Phys. Rev* **134** A549

[19] Schreiber K, Klügel T and Stedman G 2003 *J. Geophys. Res* **108** 2132

[20] Adler R 1946 *Proceedings of the IRE* **34** 351 – 357

[21] Bilger H, Stedman G, Poulton M, Rowe C, Li Z and Wells P 1993 *Instrumentation and Measurement, IEEE Transactions on* **42** 407–411

[22] Schawlow A and Townes C 1958 *Physical Review* **112** 1940–1949

[23] Chow W W, Gea-Banacloche J, Pedrotti L M, Sanders V E, Schleich W and Scully M O 1985 *Rev. Mod. Phys.* **57** 61

[24] Statz H, Dorschner T, Holtz M and Smith I 1985 Laser Handbook vol 4, ed ML Stitch and M Bass

[25] Wilkinson J 1987 *Progress in Quantum Electronics* **11** 1–103

[26] Dorschner T, Haus H, Holz M, Smith I and Statz H 1980 *Quantum Electronics, IEEE Journal of* **16** 1376–1379

[27] Bilger H R, Wells P V and Stedman G E 1994 *Appl. Opt.* **33** 7390–7396

[28] Allan D 1966 *Proceedings of the IEEE* **54** 221 – 230

[29] Allan D 1987 *Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on* **34** 647–654

[30] Ng L 1993 *On The Application of Allan Variance Method for Ring Laser Gyro Performance Characterization* Lawrence Livermore National Laboratory, report UCRL-ID-115695

[31] Longuet-Higgins M S 1950 *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences* **243** 1–35

[32] Kedar S, Longuet-Higgins M, Webb F, Graham N, Clayton E and Jones C 2008 *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science* **464** 777–793

[33] Gray B, Latimer I and Spoor S 1996 *Journal of Physics D: Applied Physics* **29** 50–56

[34] Thomson K 1978 Integrated 3 axis laser gyro *Proceedings of the seminar on laser inertial rotation sensors, San Diego, CA.*

[35] Aronowitz F and Killpatrick J 1978 *DTIC Research Report ADA051675*

[36] Levinson E, ter Horst J, Willcocks M and Marine S 1994 The next generation marine inertial navigator is here now *Position Location and Navigation Symposium, 1994., IEEE* pp 121–127

[37] Song S, Park M, Park D, Hong S, Kim D and Lee J 2007 Dithering system model identification of ring laser gyroscopes *Control, Automation and Systems, 2007. ICCAS'07. International Conference on* pp 1665–1670

[38] Ljung B 1983 Bending actuator for ring laser gyroscope US Patent 4,386,853

[39] Ljung B 1979 Piezoelectric actuator for a ring laser US Patent 4,160,184

[40] Grasso M 1995 Ring laser gyroscope with tilting mirrors US Patent 5,469,258

[41] King A 1998 *GEC review* **13** 140–149

[42] Stowell W 1974 *A precision measurement of Fresnel drag in a ring laser* Ph.D. thesis Oklahoma State Univ., Stillwater.

[43] Bilger H R and Stowell W K 1977 *Phys. Rev. A* **16** 313–319

[44] Massey G 1969 *DTIC Research Report AD0711654*

[45] Penfold C, Collins R, Tufaile A and Souche Y 2002 *Journal of Magnetism and Magnetic Materials* **242** 964–966

[46] Andrews D and King T 1994 *Journal of Physics D: Applied Physics* **27** 1815–1815

[47] Macek W 1974 Ring Laser Magnetic Bias Mirror US Patent 3,851,973

[48] Andrews D and King T 1996 *Quantum Electronics, IEEE Journal of* **32** 543–548

[49] Diels J C and Lai M 1991 *Phys. Rev. A* **44** 5898–5903

[50] Scully M O 1987 *Phys. Rev. A* **35** 452–455

[51] Hinds E A, Clendenin J E and Novick R 1978 *Phys. Rev. A* **17** 670–676

[52] Kocher C A, Clendenin J E and Novick R 1972 *Phys. Rev. Lett.* **29** 615–618

[53] Perry D 1971 *Quantum Electronics, IEEE Journal of* **7** 102–102

[54] Head C E and Head M E M 1970 *Phys. Rev. A* **2** 2244–2250

[55] Ling Y and Wu S 2005 *Review of Scientific Instruments* **76** 126107

[56] Haberland H, Konzt W and Oesterlin P 1982 *J. Phys. B: At. Mol. Phys* **15** 2969–2988

[57] Ralchenko Y, Kramida A, Reader J and (2008) N A T 2010 *NIST Atomic Spectra Database (version 3.1.5)*

[58] Siegman A E 1986 *Lasers* (University Science Books)

[59] Gordon E I and White A 2004 *Applied Physics Letters* **3** 199–201

[60] Wright D 2001 *Optimisation of Ring-laser Gyroscopes* Ph.D. thesis University of Canterbury

[61] Herziger G, Holzapfel W and Seelig W 1966 *Zeitschrift für Physik A Hadrons and Nuclei* **189** 385–400

[62] Andrews D A and King T A 1989 *J. Phys. D: Appl. Phys* **22** 1315–1320

[63] Verdeyen J 1989 *Laser electronics* (Prentice Hall)

[64] Csele M 2004 *Fundamntals of light sources and lasers* (Wiley-Interscience)

[65] Silfvast W 1996 *Laser fundamentals 2ed.* (Cambridge University Press)

[66] Svelto O 1982 *Principles of lasers* (Plenum, New York, NY)

[67] Mielenz K D and Nefflen K F 1965 *Appl. Opt.* **4** 565–567

[68] Smith P 1966 *Journal of Applied Physics* **37** 2089–2093

[69] Fork R L and Pollack M A 1965 *Phys. Rev* **139** 1408

[70] Tuchin V 1990 *Dynamic Processes in Gas-Discharge Lasers (in Russian)* (Energoatomizdat, Moscow)

[71] Knutson J W and Bennett W R 1976 *Phys. Rev. A* **13** 318–325

[72] Bennett W R 1962 *Phys. Rev.* **126** 580–593

[73] Smith P 1966 *Quantum Electronics, IEEE Journal of* **2** 77–79

[74] Patel B S, Charan S, Mallik A and Swarup P 1974 *Journal of Physics D: Applied Physics* **7** L40

[75] Smith P 1966 *Quantum Electronics, IEEE Journal of* **2** 62–68

[76] Quinn T J 1999 *Metrologia* **36** 211–244

[77] Gerstenberger D, Drobshoff A and Sheng S 1988 *Quantum Electronics, IEEE Journal of* **24** 501–502

[78] De Laeter J, Böhlke J, Dè Bievre P, Hidaka H, Peiser H, Rosman K and Taylor P 2003 *Pure Appl. Chem* **75** 683–800

[79] Press W, Teukolsky S, Vetterling W and Flannery B 2002 *Numerical recipes in C/C++* (Cambridge Univ. Press)

[80] Crowell M 1965 *Quantum Electronics, IEEE Journal of* **1** 12–20

[81] Uchida T and Ueki A 1967 *Quantum Electronics, IEEE Journal of* **3** 17–30

[82] Nash F 1967 *Quantum Electronics, IEEE Journal of* **3** 189–196

[83] Uchida T 1967 *Quantum Electronics, IEEE Journal of* **3** 7–16

[84] Lamb W E 1964 *Phys. Rev.* **134** A1429–A1450

[85] Graham R 2006 *Ring Laser Gain Media* Master's thesis University of Canterbury

[86] Weber M 2003 *Handbook of optical materials* (CRC)

[87] Schuöcker D, Reif W and Lagger H 1975 *Opt. Commun* **13** 6–12

[88] Stone J and Stejskal A 2004 *Metrologia* **41** 189–187

[89] Javan A, Bennett W R and Herriott D R 1961 *Phys. Rev. Lett.* **6** 106–110

[90] Pritsch B, Schreiber K U, Velikoseltsev A and Wells J P R 2007 *Applied Physics Letters* **91** 061115

[91] Coxeter H 1983 *Introduction to geometry* (Wiley)

[92] King G C P and Bilham R G 1973 *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences* **274** 209–217

[93] Panek L 1966 *Testing techniques for rock mechanics*

[94] Thomas A, Nadeau R and Bürgmann R 2009 *Nature* **462** 1048–1051

[95] Bitter R, Mohiuddin T and Nawrocki M 2007 *LabVIEW Advanced Programming Techniques* (CRC)

[96] Dierckx P 1975 *Journal of Computational and Applied Mathematics* **1** 165 – 184

[97] Menn N 2004 *Practical optics* (Academic Press)

[98] Brent R 2002 *Algorithms for minimization without derivatives* (Dover)

[99] Kaufman A and Levshin A 2005 *Acoustic and elastic wave fields in geophysics* (Elsevier Science Ltd)

[100] Maupin V 2007 *Advances in Geophysics* **48** 127 – 155

[101] Jaeger J, Cook N, Zimmerman R and Zimmerman R 2007 *Fundamentals of rock mechanics* (Wiley-Blackwell)

[102] Hanks T and Kanamori H 1979 *J. geophys. Res* **84** 2348–2350

[103] Okada Y 1985 *Bulletin Of The Seismological Society Of America* **75** 1135–1154

[104] Gercek H 2007 *International Journal of Rock Mechanics and Mining Sciences* **44** 1 – 13

[105] Farrell W E 1973 *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences* **274** 253–259

[106] Berger J and Lovberg R H 1970 *Science* **170** 296–303

[107] Gerard V B 1973 *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences* **274** 311–321

[108] McCarthy D and Petit G 2003 *IERS Conventions (2003), IERS Conventions Centre, Verlag des Bundesamts für Kartographie und Geodäsie, Frankfurt am Main*

[109] Vincenty T 1975 *Survey Review* **22** 88–93

[110] Clark S and Tucker R 2000 *Classical and Quantum Gravity* **17** 4125–4157

[111] Mashhoon B, Gronwald F and Lichtenegger H I M 2001 *Lect.Notes Phys* **562** 83

[112] Stedman G E, Schreiber K U and Bilger H R 2003 *Classical and Quantum Gravity* **20** 2527–2540

[113] Thomas L 1927 *Philosophical Magazine and Journal of Science* **7** 1–22

[114] Belušević R 2008 *Relativity, astrophysics and cosmology* (Vch Verlagsgesellschaft Mbh)

[115] Rindler W 1977 *Essential relativity: special, general, and cosmological* (Springer Verlag)

[116] Buchman S, Everitt C W F, Parkinson B, Turneaure J P, DeBra D, Bardas D, Bencze W, Brumley R, Gill D, Gutt G, Gwo D H, Keiser G M, Lipa J, Lockhart J, Mester J, Muhlfelder B, Taber M, Wang S, Xiao Y and Zhou P 2000 *Advances in Space Research* **25** 1177 – 1180

[117] Pascual-Sánchez J 2003 *Current Trends in Relativistic Astrophysics* 330–336

[118] Iorio L 2003 *Classical and Quantum Gravity* **20** L5

[119] Einstein A 1916 *Annalen Phys* **49** 31

[120] Krastev P G, Li B A and Worley A 2008 *Physics Letters B* **668** 1–5

[121] Arnaud N, Barsuglia M, Bizouard M, Brisson V, Cavalier F, Davier M, Hello P, Kreckelbergh S and Porter E 2004 *Astroparticle Physics* **21** 201–221

[122] Beverini N, Bradaschia C, Carelli G, DiVirgilio A, Graham R, Maccioni E, Porzio A, Solimeno S and Sorrentino F 2008 G-pisa gyrolaser *Proceedings of the 5th International Symposium on Modern Problems of Laser Physics, Novosibirsk, Russia*

[123] Velikoseltsev A, Schreiber K, Klüge T, Voigt S and Graham R 2008 Sagnac interferometry for the determination of rotations in geodesy and seismology *Proceedings of the 15th St. Petersburg International Conference on Integrated Navigation Systems, St. Petersburg, Russia*

[124] Michelson A and Morley E 1887 *Am. J. Sci.* **34** 333–345

[125] Braccini S, Casciano C, Cordero F, Corvace F, De Sanctis M, Franco R, Frasconi F, Majorana E, Paparo G, Passaquieti R *et al.* 2000 *Measurement Science and Technology* **11** 467–76

[126] Black E 2001 *American Journal of Physics* **69** 79

[127] Flaminio R and Heitmann H 1996 *Physics Letters A* **214** 112–122

[128] Acernese F, Amico P, Al-Shourbagy M, Aoudia S, Avino S, Babusci D, Ballardin G, Barillé R, Barone F, Barsotti L, Barsuglia M, Beauville F, Bizouard M A, Boccara C, Bondu F, Bosi L, Bradaschia C, Braccini S, Brillet A, Brisson V, Brocco L, Buskulic D, Calloni E, Campagna E, Cavalier F, Cavalieri R, Cella

G, Chassande-Mottin E, Corda C, Clapson A C, Cleva F, Coulon J P, Cuoco E, Dattilo V, Davier M, Rosa R D, Fiore L D, Virgilio A D, Dujardin B, Eleuteri A, Enard D, Ferrante I, Fidecaro F, Fiori I, Flaminio R, Fournier J D, Frasca S, Frasconi F, Freise A, Gammaitoni L, Gennai A, Giazotto A, Giordano G, Giordano L, Gouaty R, Grosjean D, Guidi G, Hebri S, Heitmann H, Hello P, Holloway L, Kreckelbergh S, Penna P L, Loriette V, Loupias M, Losurdo G, Mackowski J M, Majorana E, Man C N, Mantovani M, Marchesoni F, Marion F, Marque J, Martelli F, Masserot A, Mazzoni M, Milano L, Moins C, Moreau J, Morgado N, Mours B, Pai A, Palomba C, Paoletti F, Pardi S, Pasqualetti A, Passaquieti R, Passuello D, Perniola B, Piergiovanni F, Pinard L, Poggiani R, Punturo M, Puppo P, Qipiani K, Rapagnani P, Reita V, Remillieux A, Ricci F, Ricciardi I, Ruggi P, Russo G, Solimeno S, Spallicci A, Stanga R, Taddei R, Tombolato D, Tonelli M, Toncelli A, Tournefier E, Travasso F, Vajente G, Verkindt D, Vetrano F, Viceré A, Vinet J Y, Vocca H, Yvert M and Zhang Z 2006 *Classical and Quantum Gravity* **23** S91

[129] Fiori I, Braccini S, Frasconi F, Travasso F and Viceré A 2004 *Virgo Note* **VIR-NOT-PIS-1390-285**

[130] Sommerville I 1989 *Software Engineering* (Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA)

[131] Clark J, Murata M *et al.* 2001 *OASIS Committee Specification* **3**

[132] Boag S, Chamberlin D, Fernandez M, Florescu D, Robie J, Siméon J and Stefanescu M 2002 XQuery 1.0: An XML query language

[133] Berglund A, Boag S, Chamberlin D, Fernandez M, Kay M, Robie J and Simeon J 2007 XML Path Language (XPath) 2.0. W3C Recommendation, 23 January

[134] Schorlemmer D, Wyss A, Maraini S, Wiemer S and Baer M 2004 *Orfeus Newsletter* **6** 9

[135] Blackburn K, Lazzarini A, Prince T and Williams R 1999 XSIL: Extensible Scientific Interchange Language *Proc. 7th IntóÀÊl Conf. High-Performance Computing and Networking* pp 513–524

[136] Kangasharju J and Tarkoma S 2007 Benefits of alternate XML serialization formats in scientific computing *Proceedings of the 2007 workshop on Service-oriented computing performance: aspects, issues, and approaches* (ACM) p 30

[137] Kern R 2007 *NumPy discussions* URL http://svn.scipy.org/svn/numpy/trunk/doc/neps/npy-format.txt

[138] Demetriu S and Trandafira R 2003 *An. St. Univ. Ovidius Constanta* **11** 57–68

[139] Huerta-Lopez1 C, Shin Y, Powers E and Roesset J 2000 Time-frequency analysis of earthquake records (WCEE)

[140] McLeod D, King B, Stedman G, Schreiber K and Webb T 2001 *Fluctuations and Noise Letters* **1**

[141] Quinn B and Hannan E 2001 *The estimation and tracking of frequency* (Cambridge Univ Pr)

[142] Andreas Freise K S 2010 *Living Reviews in Relativity* **13** URL http://www.livingreviews.org/lrr-2010-1

[143] Basham P and Giardini D 1993 *Ann. Geofis* **36** 15–24

[144] Brohan P, Kennedy J, Harris I, Tett S and Jones P 2006 *J. Geophys. Res* **111**

[145] Saff E and Kuijlaars A 1997 *The Mathematical Intelligencer* **19** 5–11

Glossary

Alternate-split	The operation technique where split-mode is used and the order of splitting is alternated rapidly to obtain Sagnac frequency and FSR.	108
Backscatter	The scattered light from imperfections on the mirror surfaces which is injected into the co-rotating beam.	17
Clear aperture	The region in the centre of a laser mirror which is guaranteed to meet the manufacturers specification.	187
Corner boxes	The boxes which house the laser mirrors and allow them to be aligned.	10
Gain starvation	The technique of operating at a power level just above the lasing threshold in order to ensure single-mode operation.	47
Geometric scale factor	The geometric part of the scale factor in Equation 2.3, i.e. A/P	9
Heterolithic construction	Where the geometry of the laser is fixed by construction on a base made of multiple materials not especially regarded for their stability. The UG lasers are examples of this type of construction.	9

Injection seeding	When a external 'seed' laser beam is mode matched to the cavity of another laser in order to alter the distribution of the initial spontaneously emitted photons.	114
Large ring laser	The class of ring laser with perimeter of the order 4 m – 30 m. The C-II and G lasers are examples of this class of device. This class of device represents the largest practical ring laser that can be made using a monolithic construction.	10
Mode-coupling	The phenomena by which a fixed phase relationship is established between multiple running modes and causes a stable Sagnac signal.	58
Monolithic construction	Where the geometry of the laser can be regarded as fixed by construction on a base made of a single piece of high stability material. The large ring lasers G and C-II are examples of this type of construction.	9
Multi-mode threshold	The power level at which a second mode can start up and challenge a previously operating single mode for gain.	50
Ring laser	When used in this thesis, the term 'ring laser' refers to a self contained, active laser interferometer designed for measurement of inertial rotation through the Sagnac effect.	7
Sagnac contrast	The ratio of the peak-to-peak value of the Sagnac signal to the mean value.	95
Sagnac frequency	The rotation induced frequency splitting in a ring laser gyro.	8

Small ring laser	The class of ring laser with perimeter typically less than 1 m. Most small ring lasers are designed for navigation and have a requirement for non-reciprocal elements or moving parts to avoid lock-in.	30
Split Mode	The mode of operation that occurs when the longitudinal mode number is different for the clockwise and counter-clockwise beams of a ring laser.	47
Splitting order	The integer multiple of the number of mode spacings separating the operating beams, disregarding the much smaller Sagnac splitting.	47
Ultra-large ring laser	The class of ring laser with perimeter of the order or larger than 30 m. The UG series lasers are examples of this class of device. All such devices make use of hetrolithic construction.	10
UTC	Coordinated Universal Time, as defined by the international earth rotation and reference systems service (IERS).	168

List of Nomenclature

$\hbar$	The reduced Planck constant $\frac{h}{2\pi}$, sometimes called the Dirac constant.
λ	Wavelength of circulating beam
$\mathbf{\Omega}$	Rotation vector being measured
$\mathbf{A}$	Area enclosed by ring
$\mathcal{F}()$	The Fourier transform function.
$\mathfrak{Im}()$	Function returning the imaginary value of the argument.
$\mathfrak{Re}()$	Function returning the real value of the argument.
Ω_E	Earth rotation rate
Ω_q	Quantum limit to rotation rate measurement.
Ω_T	Gyroscope precession rate from Thomas precession.
Ω_{DS}	Gyroscope precession rate from the de Sitter precession.
Ω_{LT}	Gyroscope precession rate from the Lense-Thirring effect.
ϕ	Latitude
σ	Lasing transition cross section
τ_{eff}	Effective recovery time of the lasing transition
θ	Longitude
θ_S	The angle of tilt from local vertical towards the south

θ_T Angle of tilt to the north.

c The speed of light in free space

c_s Sagnac contrast, see Section 4.3.

f_0 The optical frequency of the lasing transition, value depends on transition and isotope.

f_R Sagnac reference frequency (normally set as close as possible to the nominal Sagnac frequency).

f_s Sagnac frequency

G Gravitational constant

G_0 Single pass gain

h Planks constant

i The unit imaginary.

I_{mm} The multi-mode threshold in terms of intra-cavity intensity measured at the gain tube.

I_{sat} Saturation intensity

k_b The Boltzmann constant

l_w Lorentzian line with (FWHM)

l_{w0} Natural linewidth (FWHM)

M_0 Scalar seismic moment

M_W Seismic moment magnitude

n Refractive index inside the laser cavity at the frequency of the lasing transition.

N_A Avogadro's number

P	Perimeter or path length of ring
P_i	Intra-cavity power
R	The ideal gas constant
R_E	Radius of the earth
S	Splitting order
T_{solar}	Length of the solar day.
t_r	Cavity ring down time (from intensity)
	Photodiode
	Photomultiplier tube
	Optical Shutter
	Transimpedance amplifier
FSR	Free Spectral Range

Index